International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 12, pp: 444-449 www.ijaem.net ISSN: 2395-5252

Image-Based Detection of Plant Diseases: Machine Learning to Deep Learning Journey.

Simran and Md Areeb Uddin, Mrs Vandana Chaudhary and Mrs. Namita Goyal
Department of Information Technology, Maharaja Agrasen Institute of Technology, Rohini, New Delhi.

Date of Submission: 20-11-2021 Date of Acceptance: 05-12-2021

ABSTRACT

In this project, we are predicting plant related diseases by using machine learning algorithms and image detection techniques. A considerable per cent of the tomato plant gets destroyed due to the diseases, which can severely affect the farmer's yield. To overcome the loss of farmers related to disease of the tomato plant we are proposing a system which will be able to detect the disease of the plant by analysing the symptoms related to the leaf of the plant. It will be able to detect the disease related to the plant and will give some suggestions for disease prevention and cure.

In order to detect whether the plant is suffering from any kind of diseases or not we will capture the image of the leaf. This image will be classified on the basis of our model. Multiple pre-processing will be done on the image and it will be fed to the prediction model which will tell us about the disease of the plant.

I. INTRODUCTION:

1.1. Overview:

The prevention and control of plant disease have always been widely discussed because plants are exposed to the outer environment and are highly prone to diseases. Normally, the accurate and rapid diagnosis of disease plays an important role in controlling plant disease, since useful protection measures are often implemented after correct diagnosis To identify the disease of the tomato plant, we are using image processing and machine learning algorithms which will help us in the process. We are detecting the tomato plant disease by analysing the image of the tomato plant leaves. The image of the plant leaves will be classified on the basis of our model. Multiple pre-processing will be done on the image and then it will be fed to the prediction model which will tell us about the disease of the tomato plant.

1.2. Motivation:

Major reasons for crop failure include plant diseases, improper nutrition, natural disasters etc. Crop failure leads to food scarcity, price hike and even farmer's suicide. Crop failure is the reason for 17% of the farmer's suicide cases [1]. A considerable per cent of the tomato plant gets destroyed due to the diseases, which can severely affect the farmer's yield. To overcome the loss of farmers related to disease of the tomato plant we are proposing a system which will be able to detect the disease of the plant by analysing the symptoms related to the leaf of the plant. It will be able to detect the disease related to the plant and will give some suggestions for disease prevention and cure.

1.3. Problem Statement:

India is an agricultural nation. About 70 per cent of the rural households are actively dependent on agriculture for their livelihood. Farmers have a wide range of diversity to select suitable fruit and vegetable crops. The large numbers of crops of the farmers get destroyed due to various factors such as drought, flood, hail, diseases, etc. So, to overcome the loss in production due to the diseases related to the plant. We are proposing a system which can detect the disease of the plant and can suggest some prevention and cures for that disease based on the image recognition technique.

1.4. Objective:

As we know India is an agricultural nation and its rural population is solely dependent on farming. Our proposed system can help farmers to minimize crop failure due to diseases and can help curing and preventing them. This will also prevent sudden hike in crop prices as well as reduce farmer's suicide cases. We are using image recognition techniques to identify the disease of the plant and trying to control and cure it. This system will be useful to local farmers and enthusiast gardeners, who perform the production and cultivation of tomato plants. The detection model can be used for transfer learning for other crops.

1.5. Plant diseases analysis and its symptoms:

The RGB image feature pixel counting techniques is extensively applied to agricultural science. Image analysis can be applied for the following purposes:

1. To detect plant leaf, stem, and fruit diseases.
2. To quantify affected area by disease.
3. To find the boundaries of the affected area.
4. To determine the color of the affected area
5. To determine size & shape of fruits.

Following are some common symptoms of fungal, bacterial and viral plant leaf diseases:

1.1.1. **Bacterial disease symptoms**

The disease is characterized by tiny pale green spots which soon come into view as water- soaked. The lesions enlarge and then appear as dry dead spots as shown in figure 1(a), e.g. bacterial leaf spot have brown or black water-soaked spots on the foliage, sometimes with a yellow halo, generally identical in size. Under dry conditions the spots have a speckled appearance.

1.1.2. **Viral disease symptoms**

Among all plant leaf diseases, those caused by viruses are the most difficult to diagnose. Viruses produce no tell-tale signs that can be readily observed and often easily confused with nutrient deficiencies and herbicide injury. Aphids, leafhoppers, whiteflies and cucumber beetles insects are common carriers of this disease, e.g. Mosaic Virus, Look for yellow or green stripes or spots on foliage, as shown in figure 1(b). Leaves might be wrinkled, curled and growth may be stunted.

(a) Bacterial leaf spot (b) mosaic virus

Figure 1. Bacterial and Viral disease on leaves

1.1.3. **Fungal disease symptoms**

Among all plant leaf diseases, those caused by fungus some of them are discussed below and shown in figure 2, e.g. Late blight caused by the fungus Phytophthora infesters shown in figure 2(a). It first appears on lower, older leaves like water-soaked, gray-green spots. When fungal disease matures, these spots darken and then white fungal growth forms on the undersides. Early blight is caused by the fungus Alternaria solani shown in figure 2(b). It first appears on the lower, older leaves like small brown spots with concentric rings that form a bull's eye pattern. When disease matures, it spreads outward on the leaf surface causing it to turn yellow. In downy mildew yellow to white patches on the upper surfaces of older leaves occurs. These areas are covered with white to greyish on the undersides as shown in figure 2(c).

(a) late blight (b) early blight (c) downy mildew

Figure 2. Fungal disease on leaves

II. LITERATURE REVIEW:

Literature surveys are the survey of scholarly sources on a specific topic. It provides an overview of current knowledge, allowing you to identify the similar theories, methods and the drawbacks of the existing researches. Thus, a detailed literature survey has been carried out for this project as well.

Fungi usually cause diseases that affect the plants, and they typically attack the leaves. Viral and bacterial pathogens cause many others. Precision in agriculture has improved with the increased use of ML and its related features. The reduced production quantity in agriculture hurts many people and animals, which requires modern technology to solve. The extraction and detection of diseases are easier when the image-based detection system is used because of its high accuracy and reduced complications and duplication of data. In some plants like tomatoes, the use of the images to determine the diseases that affect them and the extent of the damages cannot be achieved unless there is a high accuracy rate. . The survey on plant diseases shows that many diverse factors determine how technology-based image detection is applied. In other words, the diseases that cause visible dents and changes on the plants are the ones that can be detected using this technology as opposed to the ones that cause damages that cannot be detected from the plants' images. The analysis in this research shows that plant diseases are usually detected when they start showing an impact on the physical appearance of the plants.

The main challenge affecting the field of agriculture is the reduction in production and poor-quality production in plants. The challenge is a result of the poor detection and management of the diseases that affect the plants. The challenge is also extended to affect human beings in several ways. The reduced plant cover due to plant diseases means that global warming, famine, and reduced air purification ensue. Hyper spectral imaging has become a reliable way of detecting crop diseases on time. It is hard to determine the factors that lead to the diseases unless they are detected on time. In other words, if a disease is detected on time, it is easy to relate it to the possible factors that lead to its occurrence. For example, scientists could determine if there was a change in weather or climate that could have led to the occurrence of the disease.

Further research shows an inadequate database that could be used to provide background knowledge for comparing the images taken. The other challenge is that the symptoms and characteristics of the diseases are diverse and could be similar to a certain degree [19]. For example, many diseases could lead to the wilting of leaves. The challenge is yet to be resolved because more and new images are uploaded progressively by experts.

The other challenge is the lack of suitable instruments for use in the work of image detection. Most of the experts in the field do not have the equipment they require to analyse the images they get from the field, and this makes it hard for them to acquire accurate data and identify the diseases. The other one is that there is a low rate of implementation in some areas due to the regulations put in place to ensure the credibility and reliability of the data from these analyses. For example, after the 4th and 6th International Conference on Machine Learning and Soft Computing, there have been many regulations that may derail the use of ML in some parts. The rules discourage some of the results from the ML functions from being applied in practice because they do not meet the required parameters.

2.1 Existing System:

The used of multiple linear regression technique for image recognition and image processing. They were able to achieve the accuracy of 90% for the minor plant diseases but their system accuracy falls drastically for top accuracy of 71.32% for the major plant disease which can severely damage the crops. The accuracy of their system can improve if the dataset is improved in the terms of both quality and quantity.

Author has provided concrete evidence that the Convolution Neural Networks are very much useful for the object detection and image classification techniques. Their system can detect the plant related disease from the classes of 12 different types of plant. They are able to achieve an accuracy of about 85.53 % for the plant disease detection. Their real time accuracy is about 30% when the system is tested for the real-world images of the plant leaves.

2.2 Proposed System:

The convolution neural networks are actively used in the object detection and the image classification techniques. We are using Convolution Neural Networks (CNN) for the model training and classification in this project of tomato plant disease detection. Using the deep convolution neural network architecture, we trained our machine learning model on the images of the tomato plant leaves with the goal to identify the disease of the tomato plant by taking images with our unique model. Within the data set, there are thousands 10 of images of the tomato plant leaves which are classified into the 8 Tomato plant related diseases. Using our model, we have achieved the top accuracy of 99.35%. The model correctly classifies tomato plant disease and whether it is healthy based on the 9 classes in 993 out of the 1000 images. Our model can easily be used in smartphones, web apps and other platforms due to less requirement of CPU as the classification itself is very fast. Our model can easily identify the tomato plant related disease, if the image is contaminated i.e. the background is not homogeneous. Then also, our model can easily extract the features from the image of the tomato plant leaves and it can classify the image into various 9 classes. Our model will classify the image and will result in different possible classes and their numerical probability, that they will fall into. However, there are a number of limitations at the current stage that need to be addressed in the future work. When the model is tested on the set of images whose conditions different from the images used for training, the accuracy of our model is reduced considerably, to approximately 46.53 per cent. We can work on that limitation by extending our dataset which now consists of just 1000 images. When we will increase our dataset to consist of images for various parameters such as different lighting conditions, weather, resolution of images. The accuracy of our model will increase substantially. This will help us to identify the tomato plant disease more reliably for the real-world application. Finally, the approach presented here is not meant to replace any existing systems for the tomato plant disease detection, but this model will rather improve the performance. The tests which are done by the laboratory are ultimately always more reliable than detection basis on the tomato plant leaves images alone, and sometimes it's very challenging to detect the disease of the plant just on the basis of their visual inputs only. With the advancement in technology, the quality of images captured by the mobile devices are improving significantly, which will result in high accuracy and high certainty of prediction done by our model.

III. METHODOLOGY:

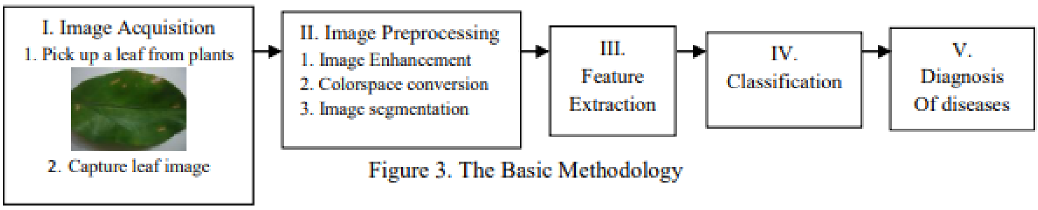


Figure 3. The Basic Methodology

3.1 Understanding data:

The dataset was provided by the kaustubh b. The dataset consists of the images of tomato plant leaves which are categorized according to the different diseases. The dataset consists of 8 folders which store the images of Tomato plant leaves suffering from different types of diseases and a folder of healthy Tomato leaves. The dataset consists of images of tomato plant leaves with the diseases such as early blight, leaf mound, late blight, etc. The dataset consists of the tomato plant leaves which are healthy. There are more than 200 images of the tomato plant leaves from each of the classes of the diseases and the healthy plant leaves.

3.2 Architecture:

Figure 4.2.1 depicts the program's flow and the different steps which are involved in model creation and prediction along with the different preprocessing methods.

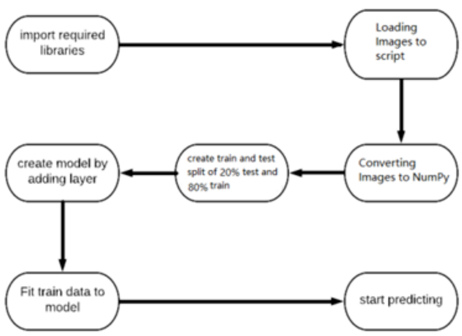
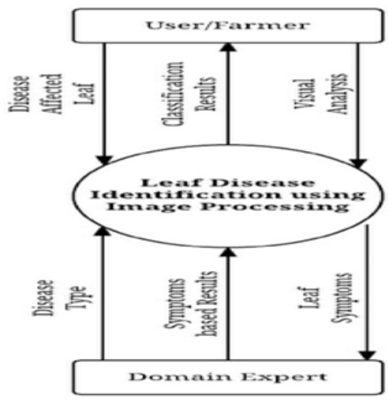


Figure 4.2.1: Model creation steps



DFD Diagram L0

3.3 Pre-processing:

In the second step, this image is pre-processed to improve the image data that suppress undesired distortions, enhances some image features important for further processing and analysis task. It includes colour space conversion, image enhancement, and image segmentation. The RGB images of leaves are converted into colour space representation. The purpose of the colour space is to facilitate the specification of colours in some standard accepted way. RGB images converted into Hue Saturation Value (HSV) colour space representation. Because RGB is for color generation and his for colour descriptor. HSV model is an ideal tool for colour perception. Hue is a colour attribute that describes pure color as perceived by an observer.

Images are resized into dimensions of 256 x 256 and loaded into the NumPy array of images. We take just the images of single leaves attached to the Tomato/cotton plants. Performed model optimization and prediction, both on the downscaled images.

We obtained three different versions of the Original Dataset by performing image operations and transformations. The versions are defined below one by one:

● In color: In this version, various transformations are applied only to the dimensional factors like angle, height and width, orientation. Keeping the colors untouched and original.

● Gray-scaled: In this version, the dimensions, straighten property(angle), and orientation is kept original. Only the colors are downscaled to grey.

● Segmented: In this version the leaves are segmented from every image of the Original Dataset i.e. the segmented data only contains the leaves only, not even the background. Segmentation can decrease the Dataset biasing and also reduces size. Thus, it also improves processing speed.

3.4 Model creation:

Tensor flow Keras library is used to get the layers from libraries and they are arranged in different order. We are creating four layers for the model creation. For the very first layer of the CNN model is defined by the initial resolution of 256 x 256.

Layers are stacked up in the following manner:

1. A convolution layer with 32 filters, a filter size of 3 and an activation function 'relu'.
2. An average pooling layer with a window size of 2.
3. A convolution layer with 32 filters, a filter size of 3 and an activation function 'relu'.
4. An average pooling layer with a window size of 2.
5. A convolution layer with 64 filters, a filter size of 3 and an activation function 'relu'.
6. An average pooling layer with a window size of 2.
7. A Flatten layer with default parameters.
8. A dense layer with 64 units and an activation function 'relu'.
9. A dropout layer with a window size of 0.5.
10. A dense layer with 9 units and an activation function 'sigmoid'.

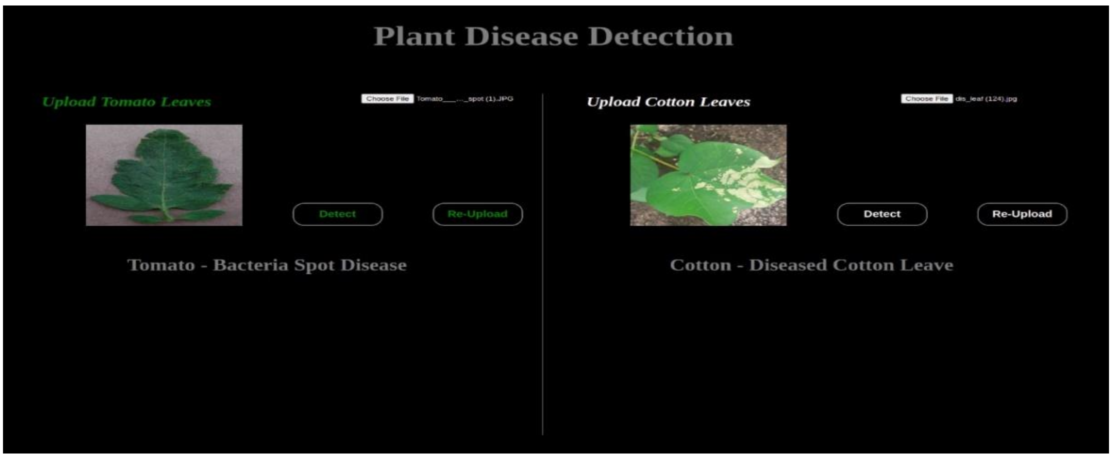
3.5 Model Training:

We trained our model with 100 epochs on Google's Data Science Community 'Kaggle'. Which provides us the resources for training the model in the best time by the use of GPUs. As the learning graph becomes constant after a certain number an epoch hence we go with 100 (not too less and not too much). The Batch size used for training is 64.

3.6 Further enhancements:

In order to improve our model performance, we can use an extensive set of data. The data which we are getting from the source has limited images, which restricts the learning ability of the system. We can use images from various different sources and can use images of different quality and in the different lighting conditions which will significantly increase our performance.

IV. RESULT:



V. CONCLUSION:

At the outset, we note that on a dataset with 9 class labels, we have achieved accuracy of 99.25 while evaluating with test and train data. However, the accuracy of prediction in the real world may vary from 70 to 100 percent (rounded off) as it also depends on the quality of images, size of images, color saturation, angle of leaf, segmentation of leaf and other properties

The system presented here is not meant to replace any existing systems for the tomato plant disease detection, but this system will rather improve the performance of the existing systems and help in upgrading them. The tests which are done by the laboratory are ultimately always more reliable than detection basis on the tomato plant leaves images alone, and sometimes it's very challenging to detect the disease of the plant just on the basis of their visual inputs only. With the advancement in technology, the quality of images captured by the mobile devices are improving significantly, which will result in high accuracy and high certainty of prediction done by our system. The improvement in image resolution will help in capturing the more details about the tomato plant leaves which eventually will lead to the better prediction and higher accuracy.

REFERENCES:

[1]. Prof.Sanjay B. Dhaygude, Mr.Nitin P. Kumbhar, Agricultural plant Leaf Disease Detection Using Image Processing, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering , S & S Publication vol. 2, Issue 1, pp: 599-602, 2013.

[2]. Guiling, Jia, Geng, and Tianyu, “Plant Diseases Recognition Based on Image Processing Technology,” Journal of Electrical and Computer Engineering, 21-May2018.

[3]. Mohanty, Sharada, D. P., Marcel, and Hughes, “Using Deep Learning for ImageBased Plant Disease Detection,” Frontiers, 06-Sep-2016.

[4]. **http://www.fao.org/india/fao-in-india/india-at-a-glance/en/**

[5]. **https://ieeexplore.ieee.org/document/8308186**

[6]. S. R. G. Reddy, G. P. S. Varma, and R. L. Davuluri, “Optimized convolutional neural network model for plant species identi fi - cation from leaf images using computer vision, ” International Journal of Speech Technology, 2021.