PLANT DISEASE DETECTION

Bhavesh Sandesh Sawant[1], Vaishnavi Vinay Pawar[2], Dhanashree Rajendra Jadhav[3], Hritik Mahendran Mayekar[4]

[1] *Student, Computer Engineering, Rajendra Mane College Of Engineering & Technology Ratnagiri, Maharashtra, India*

[2] *Student, Computer Engineering, Rajendra Mane College Of Engineering & Technology Ratnagiri, Maharashtra, India*

[3] *Student, Computer Engineering, Rajendra Mane College Of Engineering & Technology Ratnagiri, Maharashtra, India*

[4] *Student, Computer Engineering, Rajendra Mane College Of Engineering & Technology Ratnagiri, Maharashtra, India*

ABSTRACT

Early Disease Detection and pets are important for better yield and quality of crops. With a Reduction in the Quality of Agricultural Products, Disease plants can lead to huge Economic Losses to Individual farmers. In a country like India whose major Population is involved in Agriculture, Finding the disease at an early stage is very important. Faster and more precise predictions of plant disease could help reduce the losses. The Significant advancement and developments in Deep Learning have given the opportunity to improve the performance and accuracy of detection of object and recognition systems. This Paper focuses on finding plant diseases and reducing economic losses. We have proposed a deep learning-based approach for image recognition. System Proposed in the paper can Detect the different types of disease efficiently and have the ability to deal with complex scenarios. The training model achieves an accuracy of 99.35% which depicts the feasibility of a Convolution Neural Network and presents the path for AI-based Deep Learning Solutions to this Complex Problem.

Keywords: - *plant disease, late blight, early blight, deep learning, Convolution Neural Network(CNN)*

---

1. INTRODUCTION

Plants have many benefits for human life. The main human need is carbohydrates which are foods that contain carbohydrates. But in the development of crops, some obstacles must be faced by farmers, this obstacle is in the form of leaf disease. If not treated promptly, the leaf disease will result in a decrease in farm income which means a decrease in food production. So, it is necessary to detect plant disease at the right time to effectively control and prevent it. The most common diseases in plants are diseases that attack the leaves of plants, namely early blight and late blight. Cold and humid places are one of the factors for leaf disease in plants.

A leaf disease called early blight has early symptoms characterized by circular spots on the middle of the leaves and it could also be on the edges of the leaves as shown in Fig. 1 (b). Then these spots will widen and the color of the leaves turns brown, the fungus Alternia solani is the cause of this leaf disease. Furthermore, Microbe Phytopthora infestans de Bary is the main cause of leaf disease called late blight; plant leaves affected by this disease can cause plant damage. Fig. 1 (a) shows a leaf with late blight, marked by the appearance of black lesions on the leaves, and will continue to propagate diseases in plants are diseases that attack the leaves of plants, namely early blight and late blight. Cold and humid places are one of the factors for leaf disease in plants.

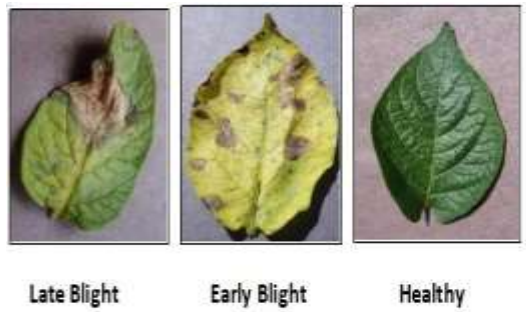


Fig 1: disease (a) Late blight (b) Early blight

This identification can help agricultural managers to provide effective and efficient handling of plants that is not healthy or abnormal. With the development of technology today there have been many digital image studies in agriculture both to identify diseases and identify good agricultural production. One of these digital image studies is to identify leaf rot in plants.

This study aims to create a system that can help farmers or agricultural managers in identifying diseases in leaves by using data on leaf images. The identification of leaves in plants is divided into three parts, namely plants with healthy leaves, late blight, and early blight. So, this study will identify this using the Convolutional Neural Network (CNN) architecture which is one of the Deep Learning methods. The data used in the form of disease data on leaves of plants were obtained from the website Kaggle with the name Plant Village

2. RELATED LITERATURE

Previous studies that have conducted trials in detecting diseases in plants include the following:

Suttapakti in his research titled " Leaf Disease Classification Based on Distinct Colour and Texture Feature Extraction" in identifying leaf diseases in plants focuses on feature extraction because according to him doing classification will get good results depending on accurate extraction features. He was comparing the color extraction feature between Colour Moments (CM) and Boundary Colour (BC) and Maximum-minimum Colour Difference (MCD) on 300 leaf images and the results that MCD has better accuracy than others witha level of 82.5%.

Islam. M in his research "Detection of Diseases Using Image Segmentation and Multiclass Support Vector Machine" uses the Support Vector Machine (SVM) method to detect disease in leaves with the amount of leaf image data used 200 data of diseased leaves and 100 normal or healthyleaves. The process in this study is to segment the image to display only leaf disease without displaying the background and normal leaves in the sense of leaf green, then the leaf image will be extracted using the Gray Level Co-occurrence Matrix (GLCM). Theresults of this study have an accuracy of 95%.

Research conducted by Mim "Leaves Diseases Detection of Tomato Using Image Processing" conducted a classification with 6 classes in the form of healthy, late blight, yellow curved, tomato mosaic, bacterial spots, and sectorial leaf spots. The algorithm used for classification is Convolutional Neural Network (CNN). In this study, the best accuracy results were obtained by running 30 epochs with 92.61% accuracy training results and validation accuracy produces 96.55%.

Prakash in his research entitled "Detection of Leaf Diseases and Classification using Digital Image Processing" segmented to remove background in leaf images using K-Means clustering, and also extracted leaf textures using the Gray Level Co-Occurrence Matrix (GLCM) feature so that showing only leaves that are not green or that are not normal. In the final stage, classification using the Support Vector Machine(SVM) algorithm produces an accuracy of 90% of the image data used in as many as 60 images.

From the description of the research that has been done in identifying plant diseases using leaf image datasets, it can be seen that there are those using the Support Vector Machine (SVM) algorithm with additional features extracting leaf texture features and producing good accuracy. The research conducted by Mim is to detect diseases of tomatoes with Convolutional Neural Network (CNN) and provide excellent accuracy in 6 classes. So, this study will identify diseases in plants using theCNN architecture from Deep Learning Method.

3. PROPOSED METHOD

In this paper, there are several stages of research completion as shown in Fig 1 in the form of a research framework. In this research framework, there are four stages in the form of dataset collection, image data pre-processing, training data, and data evaluation.

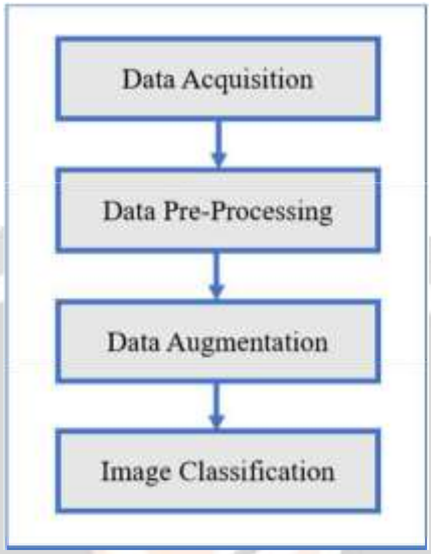


Fig -2: Research Framework

3.1 Data Collection

The dataset used in this study is a picture of the leaves of a plant divided into three classes: the healthy leaves shown in Figure 2, the early blight shown in Figure 3, and the late blight shown in Figure 3. This dataset was obtained from the Kaggle website under the name "Plant Village Dataset" uploaded by Arjun Tejaswi and the latest changes were made in October 2021. The amount of data used is 500 data late blight, 500 initial early blight data, and 150 healthy leaf data. Details of the data are agreed upon in Table 1.

Table -1: DETAIL DATASET

Samples	**Number**	**Repository**
Late Blight	1000	Kaggle (Plant Village)
Early Blight	1000	
Leaf Healthy	152	
Total	2152	

All images used in this study will be resized to 150x150 to speed up processing. Fig 3 are examples of pictures from each data class used Fig 3 (a) leaf healthy, Fig 3 (b) early blight, and Fig 3 (c) late blight.

(A) (B) (C)

Fig -3: Image data (A) leaf healthy (B) Early blight (C) Late blight

3.2 Preprocessing Data

At this stage the amount of data used is 2152 images from 3 classes as in Figure 5, the class divisions are late blight, early blight, and leaf healthy. In Table 2, the details of the distribution of each data from each class used are divided into training data and testing data with a data division of 80:20. The results of each of these data sharing willbe compared to the results of its accuracy to determine which is better in dividing the proportion of data. The leaf image used first will be resized to 150x150 to speed up classification processing.

Table -2: DETAIL OF DATASET

Dataset	**80:20**	
	Train	**Val**
Late Blight	800	200
Early Blight	800	200
Leaf Healthy	122	30
Total	1722	230

division 80:20, for data dividing 90:10, is not used in this study because considering the 152 leaf healthy data, if the data validation data used healthy leaves only 10% it will not suffice for the validation process with the number of batch sizes used in this study is 20 batch sizes.

3.3 Convolution

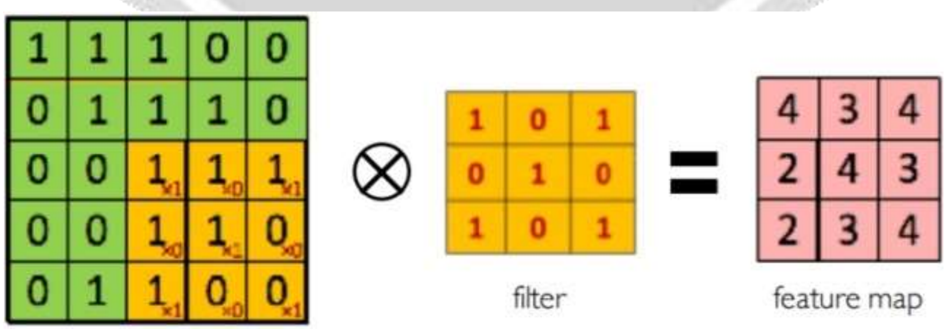


Fig 4 - Convolution Operation

The convolution process and matrix multiplication will be carried out on the filter and leaf image area. As shown in Fig. 4 this convolution process multiplies the pixels in the image by the filter pixels. This research will use 4 layers of convolution**.**

3.4 Pooling

After the convolution is carried out, then it will do pooling. The pooling that is often used is Max Pooling. Pooling here means the process carried out to get images with smaller pixels but still by maintaining the information in the image. The pooling process can be seen in Fig 5, where the image area with aparticular pixel area will be done pooling by selecting one of the highest pixels. This process is constructive because it will reduce the size of each image and will be able to speed up the classification process.

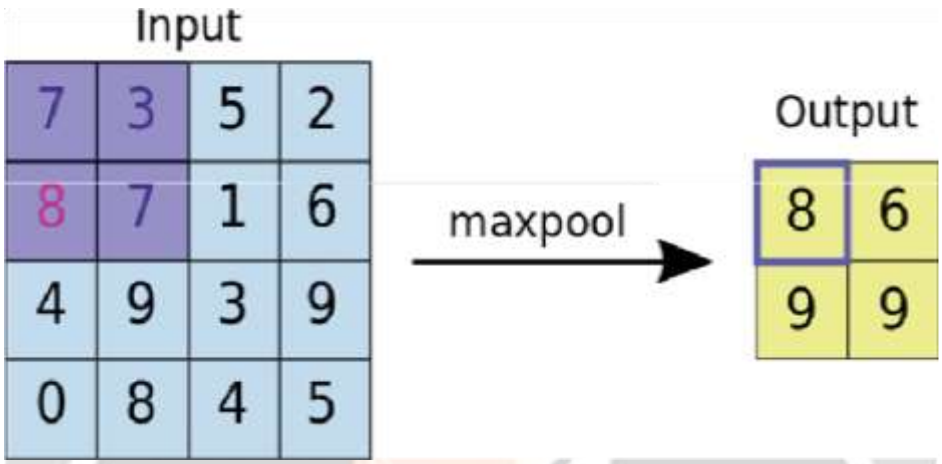


Fig 5 – Max Pooling

1.4 Classification

The next step is to classify images using the Convolutional Neural Network (CNN) architecture from Deep Learning Method. Convolutional Neural Network (CNN) architecture from Deep Learning Method is included in the supervised learning method where identification of an image by training existing image data and targeting image variables.

The convolutional layer in the Convolutional Neural Network (CNN) architecture helps neural networks in the CNN method recognize leaves based on the attributes they have. The neural network can recognize images of leaves based on the pixels in the picture

This research will use an image with a size of 150x150x 3, which means there is a 150x150 size image and this image has three channels, namely red, green, and blue (RGB). This leaf image will be convoluted first with a filter. Then pooling will be done to reduce the image resolution while maintaining image quality, pooling used is Max Pooling on the input image.

The next process is a fully connected layer, wherein this process is flattened. The purpose of flattening here is to change the feature map resulting from pooling into vector form. For more details, shown in Fig 6 architecture of the Convolutional Neural Network (CNN).

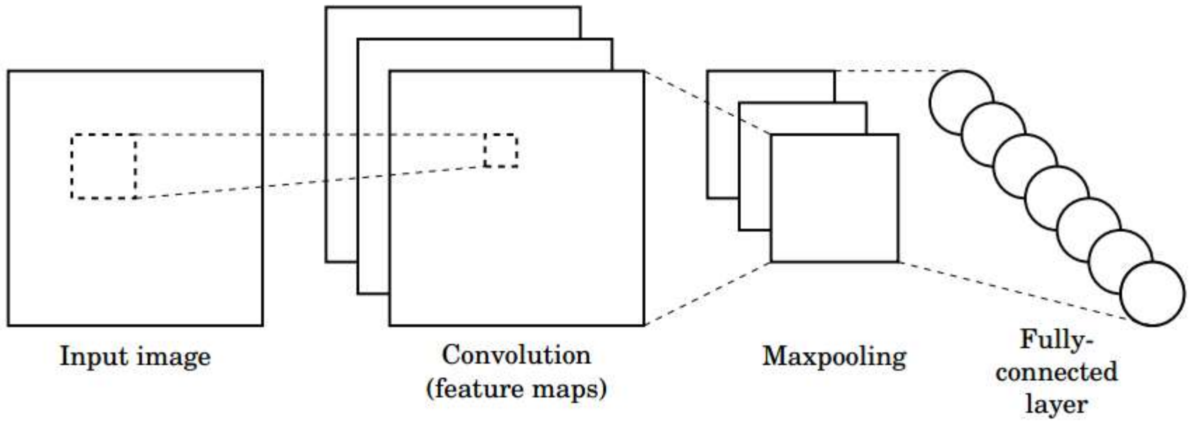


Fig - 6: Convolution Neuron Network

4. RESULT AND EVALUATION

Epoch is a training process on neural networksuntil it returns to the initial stage in one round whenall datasets go through this process. In training data with a neural network model, if you only use one epoch, this will be too large and will stress the training process in the dataset, because the data used is quite a lot, it is necessary to divide the data rate per batch (batch size). In this research, there were 32 batch sizes and to determine the number of epochs,the researchers adjusted the number of batch sizes to the number of samples used.

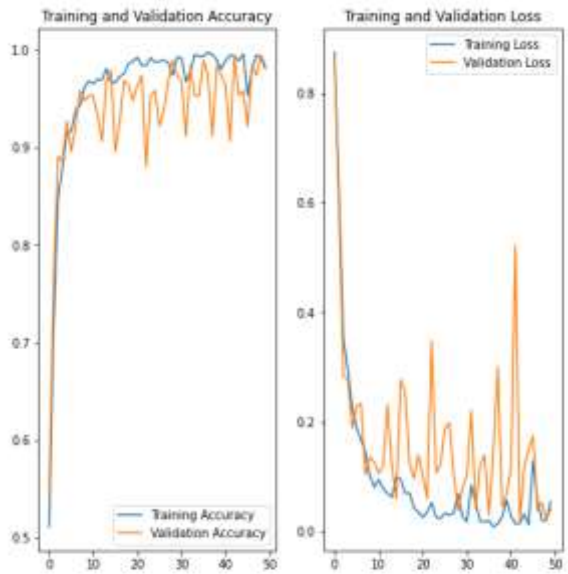
The next step is to conduct training on the leaf image which has been divided by the fit model. Table 5 shows the results of the fit model in the 80:20 that the accuracy value on the train data and the accuracy value on the validation data have increased.

TABLE – 5: RESULT FROM FIT MODEL 80:20

Epoch	Data Training		Data Testing	
	Accuracy	Loss	Value Accuracy	Value Loss
1	0.4412	0.6077	0.4364	0.5796
2	0.6522	0.4482	0.7909	0.3902
3	0.8005	0.3083	0.8091	0.3253
8	0.9258	0.1222	0.9182	0.1352
9	0.9629	0.0610	0.9636	0.0861
10	0.9655	0.0623	0.9273	0.1755

Then in Table 5 shows the results of the classification of the data divided into 80:20.

The results of the first epoch display the accuracy value on the train data as 44% with a loss value obtained that is 60% and so on until the 10th epoch displays an accuracy value of 96% with a loss value obtained at 0.06%. Whereas in the testing data, the accuracy value on the first epoch is 43% with a loss valueobtained of 57%, and so on until the 10th epoch displays an accuracy value of 92% with a loss valueobtained is 17%.



(A) (B)

Chart 1 - Training and validation from 80:20 data dividing (A) accuracy (B) loss

The graph shown in Fig 6 (a) is a graph of accuracy and the graph shown in Fig 6 (b) is a graph of loss for the 80:20 data division. This graph also shows that the fit model made with 80:20 data dividing is good because the increased accuracy and decreased loss for each epoch are as stable.

From this research it can be concluded that based on the implementation of the model, the distribution of the dataset has little effect on the results of accuracy, from the results carried out on leaf data has a slightly better result by the dataset at 80:20. The image used is changed to 150x150 size then classification is done with 50 epochs with a batch size of 32 with a total data of 2152 images. The accuracy value for the training data is 96% and for the validation, accuracy is 92% for the 80:20 data dividing.

5. CONCLUSION

From this research it can be concluded that based on the implementation of the model carried out, the distribution of the dataset has an effect on the results of accuracy, from the results carried out on leaf data has good results by dividing the dataset by 80% and 20% and the image used is changed in size 150x150. On the 50th epoch with a batch size of 32 with a total training data of 800 images and testing data of 200 images, the accuracy value on the traindata is 97% and for validation, accuracy is 92%.

6. REFERENCES

1. Lili Li, Shujuan Zhang, And Bin Wang" Plant Disease Detection and Classification by Deep Learning—", April 8, 2021

2. An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques Ms. Kiran R. Gavhale1, Prof. Ujwalla Gawa