Detection of Diseases in Plant Leaf Using CNN Technique

B. Priya Esther[a], J.Selvin Paul Peter[b], Shivam Sunil Bhosale[c], Shubham Pandey[c]

[a]Asst. Professor, Dept. of Mechatronics, SRMIST, Chennai, Tamil Nadu
*[b]Associate. Professor, Dept. of CS Engineering, SRMIST, Chennai,Tamil Nadu.**
[c]Student, Dept. of CS Engineering, SRMIST, Chennai, Tamil Nadu

*** Corresponding author :J Selvin Paul Peter.**
Tel.+919986563360
E-mail address: selwyn2009@yahoo.com

ABSTRACT
In our research, we have used Convolution Neural Networks (CNN) to detect and identify the type of leaf as well as the disease it has been affected with. The image dataset we used for the training purpose is titled 'Plant Village'. In this, the plant species were properly differentiated with respect to their species and disease they have been affected with. This image data was first made compatible with our CNN model by reducing its dimension to 227X227 pixels. The model was trained on various CNN layers to ensure that the features are extracted suitably. To deploy the model, we used stream lit web application python module. Here we could upload the leaf image and for our model to make a prediction. With our model, we were able to achieve an accuracy of above98%.

Key-Words: *Convolution Neural Network, Dataset, Accuracy, Validation and Training.*

1. INTRODUCTION

As we all know that India is a well-known as an agricultural country as the majority of the population relies on agriculture. The agricultural practices serve a dual purpose of feeding the ever-growing population and also is looked upon as a solution to the global warming crisis. Agricultural research is aimed at increasing production and food quality with reduced expenditure and higher benefit, which has gained significance in recent times. In the agricultural industry, the classification and identification of crop diseases is of great technical and economic significance. Plant diseases are unavoidable, and therefore early detection of diseases will be advantageous foragriculture. The plant disease could be infectious and the agent causing it can be viral, fungal or bacterial. Plant disease is one of all the crucial causes that reduces amount and degrades quality of the agricultural manufacture.

The plants diseases can be classified into 2 distinctive parts, namely infectious and non-infectious. The majority of plant diseases are caused by living agents which are pathogens like fungi, bacteria, viruses, parasites and nematodes and therefore fall under the category of infectious diseases. On the other hand, the non-infectious diseases are the once which are caused by abiotic agents such as different harsh environmental conditions.

Leaf is an integral part of a plant which helps it to produce energy required for its growth and nourishment. It carries out all the important processes such as photosynthesis, healthy growth of any plant can be attributed to the quality of the leaf it produces. The Leaf is also one of the most exposed part of a plant, making it prone and vulnerable to infectious diseases.

The new age of farming is expected to solve the common problems to ensure maximum yield. In any normal scenario to identify the disease it would take years of expertise in the field or the farmer would have to send the samples to the laboratory for manual testing which can take up a lot of time allowing the diseases to spread across to other neighboring plants. Early detection of such infectious diseases can save a lot of time and money to the farmer. Once the farmer understand the exact pathogen which is causing the diseases, they can use pathogen specific insecticides and get rid of the infections at an early stage.

2. LITERATURE SURVEY

In the paper published by SrdjanSladojevic, Marko Arsenovic, AndrasAnderla, DubravkoCulibrk, and DarkoStefanovic, the authors have used CNN as a technique to classify between a variety of plant leaf diseases. The model developed by the authors is an integration of ANN and CNN, developed using the CaffeNet framework

developed by the developed by Berkeley AI Research (BAIR) and by community contributors. Images which were used for the dataset were images resized to 256 × 256 to reduce the time of training. Using the above model, the authors were able to achieve an accuracy of 96.3%, after 100th training, with fine-tuning and 95.8% without fine-tuning. [1]

The Authors in the third paper talked about detecting plant diseases using support vector machines (SVM) and particle swarm optimization. The proposed system uses the initial image and enhances it using K-means clustering and then the features are extracted and the disease is detected using SVMand PSO method. The authors were able to achieve an accuracy of about 65-72% using just SVM and about 95% when SVM was integrated with PSO. [3]

The Authors of the second paper have worked on different ML algorithms to predict a plant disease. The work is based on the morphological features of the plant leaf and the techniques represented in the paper were decision tree, Naive Bayes theorem, artificial neural network and k-mean clustering and random forest algorithms. Disease development depends on three conditions-host plants susceptible to disease, favorable environment and viable pathogen. The presence of all three conditions is must for a disease to occur. [2]

The Authors in this paper talked about detecting plant diseases using the simple method of segmentation and then comparison. In this paper presented by them, the color transform of the RGB image has been used for better breakdown of the diseased spot in the plant. While in the comparison part of the process techniques like YCbCr model, HIS color Model and CIELAB color model are used. Each of the above-mentioned methods were tested and in the end CIELAB color model was used since it was capable of reducing the noise from the RGB picture. [4]

The authors in this paper describe the techniques used to detect the plant disease using image processing and describes the process of image acquisition, preprocessing, segmentation and feature extraction. In this paper spatial filter, K-means, GLCM, and SVM are discussed, and gives the implementation result of plant disease detection. Here, only pomegranate plant is used and an accuracy of 90% is achieved. [5]

In this paper the proposed system helps in detection of plant diseases and provides remedies that can be used against the disease. In the following paper Convolution Neural Network (CNN) is used to predict after the data is trained in the classifier. In this paper the software give the name and spices of the plant, confidence level and the remedy as the output, the accuracy of this is 78% as mentioned in the paper. [6]

The proposed system in the [7] is the application of K-Means clustering neural network and fuzzy logic, and then that has been composed for the clustering and classification of diseases that affect the plant leaves. Here the classification is done only for five diseases. Here the main objective is to detect the disease early and accurately and also increase the production, productivity and food quality at a reduced expense.

In the paper [8] presented by Barbedo JGA, a survey on methods that employs digital image processing for detecting, quantifying and classifying plant diseases from digital images. Any part of the plant can be infected, but in this paper that has only considered the leaves and the stem and left out the other part of the plants such as root and seeds in order to reduce the size of the paper. There are three major parts in the paper which are detection, quantification and classification and these three and further divided into several parts depending on the need and algorithm.

In [9], the authors have used advanced health monitoring sensors, which could be used to detect plant diseases with a high accuracy. The technologies included spectroscopic and image-based and volatile profiling based disease detection methods.

In [10], the authors have used machine vision and fuzzy logic, the authors of this paper have stated that the current technique is based on the manual approach of the farmers, which can be a tedious task in identifying and grading the scale of the infected plant. This manual is time-consuming and also not possible in the upcoming future.

Taking all this into consideration, the authors have gone forward with an image processing based approach which identifies and grades the disease on the leaf of the plant with the help of Fuzzy Logic.

Table 1: Accuracy for different algorithms

S.no	Machine Learning Algorithm	Accuracy
1	Decision Tree	89%
2	K-Means Clustering	83%
3	Naive Bayes	86%
4	Random Forest	83%

Table 1 above lists the accuracy of various detection algorithms

It consists of 5 modules i.e. Image Acquisition (ii) Image Preprocessing (iii) Color Image Segmentation (iv) Calculating At and Ad (v) Disease grading by Fuzzy logic.

3. PROPOSED SYSTEM

Here we proposed to use the Artificial Neural Network technique, to classify whether the plant is infected or not based on different parameters. Convolutional Neural Network (CNN) would be used to classify between the images thereby allowing the user to quickly identify the type of the disease if the plant is infected with one, the model will be trained on different datasets in the form of images to make the necessary predictions. In deep learning, Convolutional neural networks (CNNs) are one of the main categories which participate in images recognition and images classifications. CNN is widely used in objects detections, recognition of faces etc. Predominantly Python libraries such as Tensor Flow, Keras, etc. would be used in the model, the transfer learning method will also help to size down the production process and increase the model accuracy.

4. DATASET USED

The model is built using a plant leaf image dataset which had differentclasses is listed in table 2

Table 2: Dataset used in the model training

Sr.no.	Class	Number of Images in Dataset
1	Pepper Bell Leaf Healthy	1478
2	Pepper Bell Bacterial Spot	997
3	Potato Healthy	152
4	Potato Early Blight	1000
5	Potato Late Blight	1000
6	Tomato Healthy	1591
7	Tomato Bacterial Spot	2127
8	Tomato Early Blight	1000
9	Tomato Late Blight	1909
10	Tomato Leaf Mold	952
11	Tomato Septoria Leaf Spot	1771
12	Tomato Spider Mites	1676
13	Tomato Target spot	1404
14	Tomato Mosaic Virus	373
15	Tomato Yellow Leaf Curl Virus	3209

5. METHODOLOGY

In CNN, the image is taken in the form of a pixel and a random filter is generated which is then multiplied with the corresponding pixel and its neighboring pixels. A summation of these is taken and a second matrix is generated. In this, MaxPooling2D is performed and the matrix is divided into smaller divisions, the maximum pixel value of each of these smaller divisions is selected which then enhances the image. Data preprocessing is carried out in the first step of CNN and augmentations such as image resizing, reshaping is done. The image is converted in a form of an array which holds the pixel's values. Then the image is processed through various layers which are encompassed in CNN which are the convolution2D layer, maxpooling2D layers, Batch Normalization, flatten layer, dense layer

and the dropout Layer. After the model is trained successfully it can accurately classify the disease which the plant has been infected with. Shown in Fig 1 is the sequence diagram of the complete process

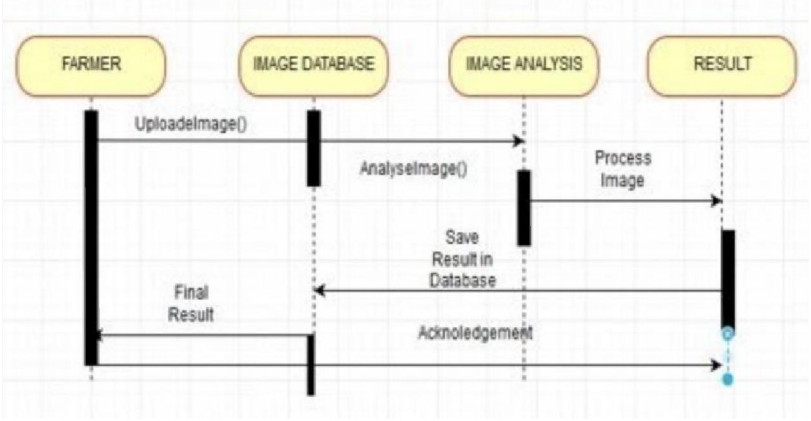

Fig 1: Sequence Diagram

6. MODULES

There are six essential and critical modules in this plant disease and detection project which are an integral part of the system. The architectural diagram is shown in fig 2

1. **Acquisition/Extraction of an image**: Acquisition of images is the step in which the image of the leaf is taken as input. The image can be captured using any camera device with focus on the leaf of the plant.
2. **Preprocessing Image:** The aim of preprocessing is to reduce the size of the image input so that it can fit in the model. The model requires the image to be of size 227X227, this can be done using any imaging cropping tool. If the model is fed with the image whose dimensions vary the mentioned size it may result in inaccuracy while extracting features from it using Convolution Neural Networks in the next step.
3. **Feature Extraction:** In the feature extraction features are extracted that can be used to determine the meaning of a given sample. This step is done with the help of Convolution Neural Network(CNN). Features such as yellow spots, rust, black dots which are an anomaly on the leaf can be found out by CNN.
4. **Disease Comparison**: In this step the extracted features from the previous module are compared with the diseases present in the model. The model has a dictionary of trained images of leafs along with their respective labels.
5. **Disease Detection:** After comparing with all the different diseases, the one with the highest accuracy is selected as the "detected disease" in the plant and is said to the output. Since we would be using the softmax function in this step, the outcome would be predicted with a score and the score closest to the nearest whole number would be selected as the output label.
6. **Final Output**: In the last step, the final output which is the detected disease is displayed along with the remedies or cure for the same. The whole process is quick and therefore would save a lot of time. Upon knowing the disease, the farmer can quickly act on the remedies to cure the infected leaf.

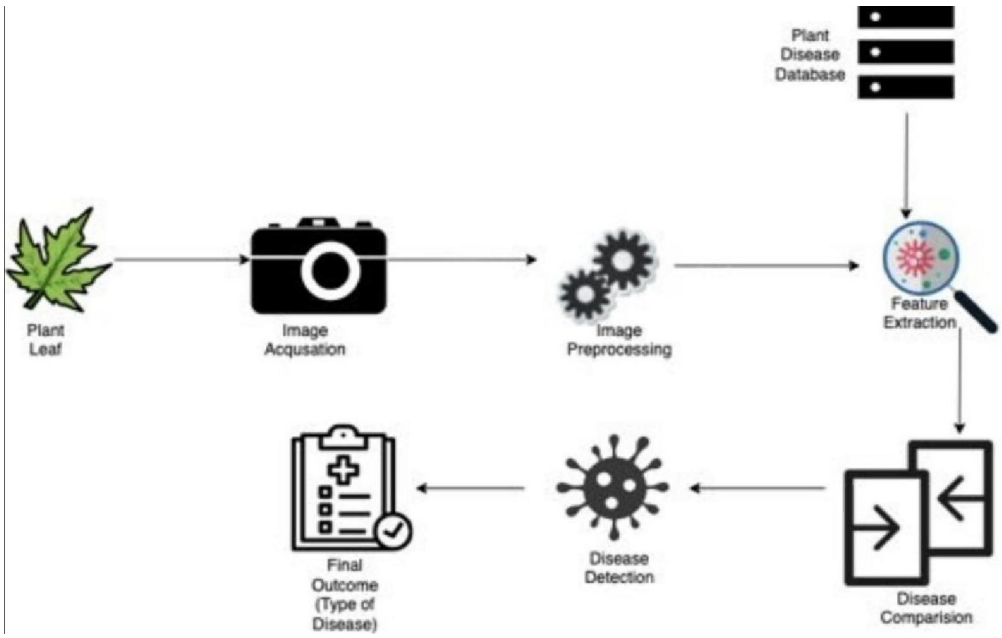

Fig 2: Architectural Diagram

By using the softmax function, the model classify the different leaf diseases

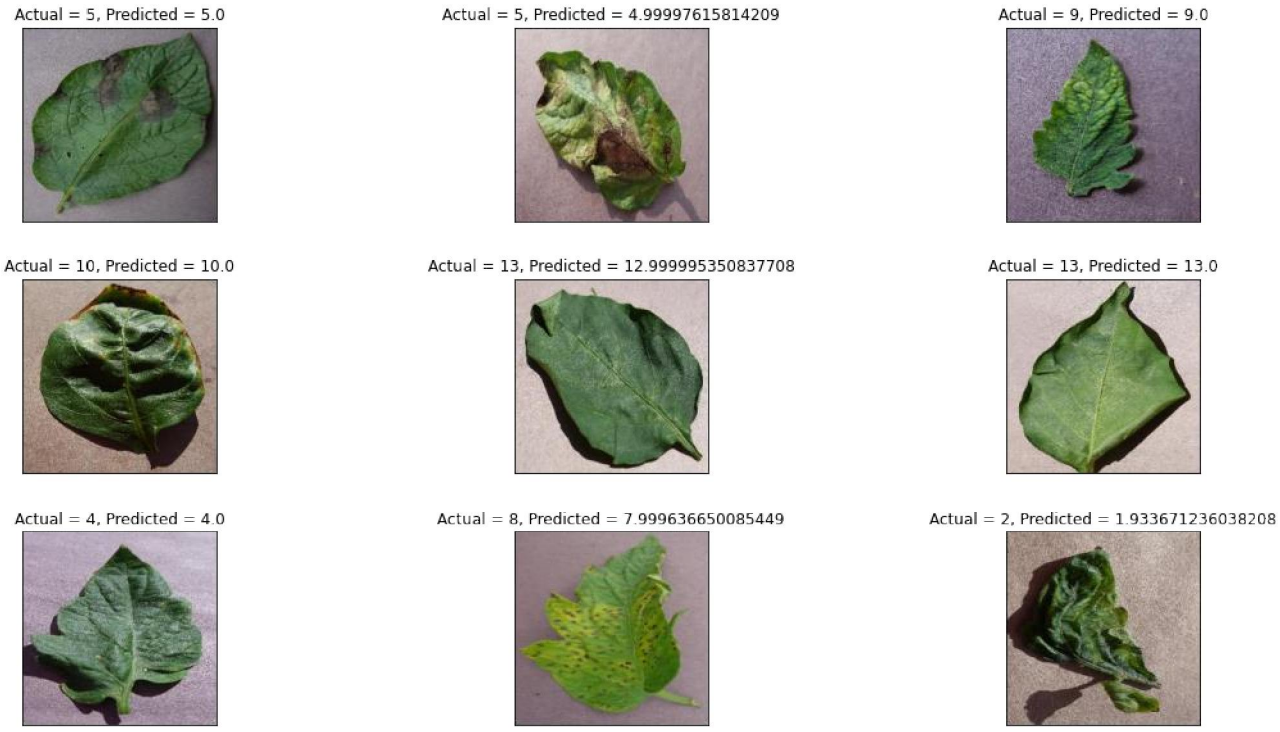

Fig 3: Prediction Made by the model

From, Fig 3 it can be said that the model had successfully assigned a label to the image given as an input. Then the model was deployed in a web application to increase the user compatibility with an easy to understand user interface.is shown in Fig 4 and Fig 5.

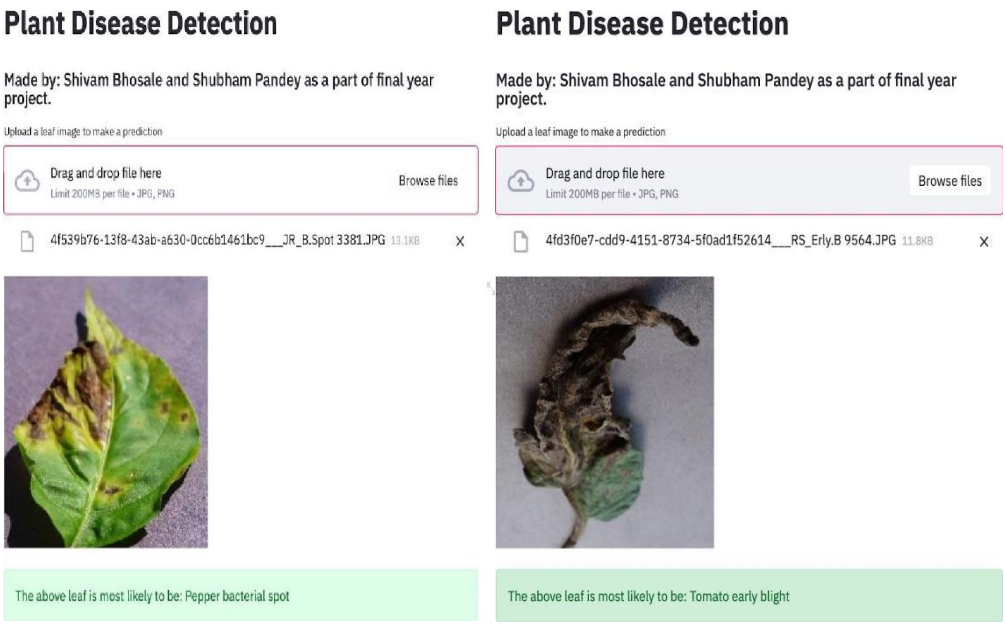

Fig 4 & 5: Deployed web application

7. Result and Analysis

During the model training the accuracies achieved at different epochs are given below table 3

Table 3: Model accuracy

Sr. no	Epochs	Accuracy
1	25	91.78%
2	50	94.89%
3	75	97.76%
4	100	98.91%

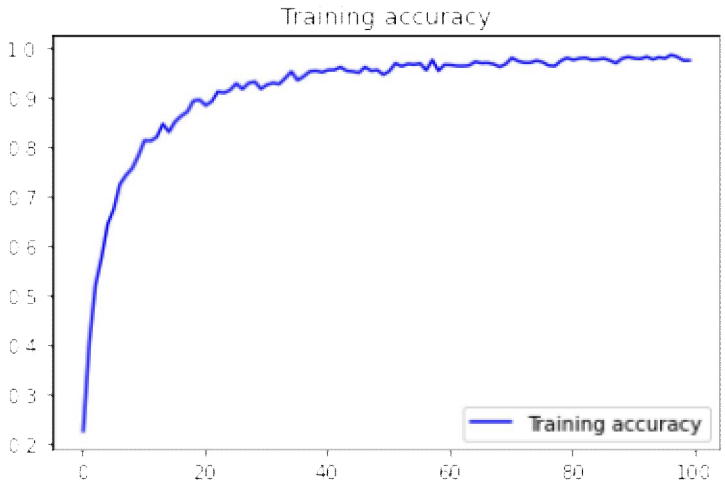

Fig 6: Training Accuracy

From Fig 6 we infer that the Training accuracy increases as the no increases and from fig 7we were able to show that we can reduce the loss function substantially.

Fig 7: Training Accuracy

8. CONCLUSION

In this paper, we have used Convolution Neural Networks (CNN) to detect and identify the type of leaf as well as the disease it has been affected with. The image dataset we used for the training purpose is titled 'PlantVillage'. The model was trained on various CNN layers to ensure that the features are extracted suitably. To deploy the model, we used streamlit web application python module. With our model, we were able to achieve an accuracy of above 98%. The graph presented in the result and analysis section prove that the model is working with efficient success.

9. REFERENCES

[1] SrdjanSladojevic, Marko ArsenovicAndrasAnderla, DubravkoCulibrk, and DarkoStefanovic. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification, Hindawi Publishing Corporation Vol. 2016 Article ID 3289801,2016.

[2] G. Prem Rishi Kranth M. HemaLalithaLaharikaBasava Anjali Mathur. Plant Disease Prediction Using ML Algorithms, International Journal of Computer Applications Vol. 182 No. 25, 2018.

[3] Prabhjeet Kaur, Sanjay Singla, Sukhdeep Singh. Detection and classification of leaf diseases using integrated approach of support vector machine and particle swarm optimization, International Journal of Advanced and Applied Science 4(8) pg. 79-83,2017.

[4] Piyush Chaudhary, Anand K. Chaudhari, Dr. A. N. Cheeran and ShardaGodaraColor Transform Based Approach for Disease Spot Detection on Plant Leaf, International Journal of Computer Science and Telecommunication Vol. 3 Issue 6,2012.

[5] Adnan Mushtaq Ali Karol, DrushtiGulhane, TejalChandiwade. Plant Disease Detection using CNN &Remedy, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 8 Issue 3,2019.

[6] PritiBadar, Suchitra.C. Disease Detection of Pomegranate Plant using Image Processing, International Journal of Computer Science and Mobile Computing Vol.5 Issue 5 pg. 328-334, 2017.

[7] Mrunalini R. Badnakhe and Prashant R. Deshmukh. An Application of K-Means Clustering and Artificial Intelligence in Pattern Recognition for Crop Diseases, International Conference on Advancements in Information Technology with Workshop of ICBMG Vol. 20,2011.

[8] Barbedo JGA. Digital image processing techniques for detecting, quantifying andclassifying plant diseases, Springer Plus volume 2, Article number: 660,2013.

[9] Sankaran S, Mishra A, Ehsani R, and Davis C. A review of advanced techniques for detecting plant diseases, Computer and Electronics in Agriculture Vol. 72 No. 1,2010.

[10] Sanjeev S Sannakki, Vijay S Rajpurohit, V B Nargund, Arun Kumar R, Prema S Yallur. Leaf Disease Grading by Machine Vision and Fuzzy Logic. Vol2(5)1709-1716,2011