PLANT LEAF DISEASE DETECTION

Kunj A. Gala
KJSCE IT 1814021
Mumbai, India
kunj.gala@somaiya.edu

Virag R. Jain
KJSCE IT 1814026
Mumbai, India
virag.j@somaiya.edu

Jai A. Mehta
KJSCE IT 1814036
Mumbai, India
jai.mehta@somaiya.edu

Vishant K. Mehta
KJSCE IT 1814038
Mumbai, India
vishant.m@somaiya.edu

Pankti K. Nanavati
KJSCE IT 1814045
Mumbai, India
pankti.n@somaiya.edu

***Abstract*—Agriculture is one of the most important sources of income for people in many countries. However, plant disease issues influence many farmers, as diseases in plants often naturally occur. If proper care is not taken, diseases can have hazardous effects on plants and influence the product quality, quantity or productivity. Therefore, the detection and prevention of plant diseases are serious concerns and should be considered to increase productivity. An effective detection and identification technology can be beneficial for monitoring plant diseases. Generally, the leaves of plants show the first signs of plant disease, and most diseases can be detected from the symptoms that appear on the leaves.**

***Index Terms*—Agriculture, Disease, Detection, Prevention, Productivity, Identification**

I. INTRODUCTION

Many people in the world depend on agriculture for their income. Indeed, we all depend on the agricultural industry directly or indirectly. When plants are affected by diseases, they have considerable negative influences on the quantity and quality of products. The risk of food insecurity increases if these diseases are not detected and diagnosed in time. Climate change and pollution are reasons for the plants to come in contact with different diseases. Cultivators in any country face severe losses because of these diseases. For leaf disease, the amount of crop production is decreasing day by day. Use of Deep learning is penetrating into every next industry one could imagine. One of them is agriculture. The main idea is the detection of early plant diseases to improve the crop yield. However, so far, the most important method to diagnosing plant diseases has been direct and visual monitoring by experienced people and plant specialists, and this method requires continuous monitoring by experts, and it is obvious that this has cost a lot of money for manufacturers. In addition, meeting agricultural specialists is not possible at all times. Especially in developing countries, farmers have to travel long distances to access specialists in agriculture and plant diseases, which in addition to spending time requires great investment. Due to the advancement of science and the introduction of new techniques, the previous methods are useless and expensive. If we can detect disease at the primary stage it will help to the growth of the production. Ignorance of this cannot be affordable and to prevent it use of chemical pesticides is very harmful. Use of pesticides damages the soil and environment. Early identification of diseases in crops is very effective and helpful for the process. With the help of Artificial intelligence and computer vision, farmers can automatically detect the plant diseases through raw images of plant leaves. Techniques of image processing are generally used in agriculture and it is applied for the detection and recognition of weeds, fruit-grading, identifying and calculating disease infestations of plants, and plant genomics. Currently, the introduction of deep learning methods turns out to be popular. Deep learning a branch of artificial intelligence is an advanced method of ML that uses neural networks that works like the human brain. General methods are like classification methods.

II. MAIN IDEA OF THE PAPER

A. Paper I - Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures

This paper mainly focuses on the detection and identification of leaf diseases in the agricultural field and on increasing the quality and quantity of the production rate. Tomato and potato which are an important constituent of our daily diet are the two crop leaves that have been considered to detect and identify diseases. Any disease which is naturally created can have serious effects on grains and vegetables. It can ultimately reduce productivity, quality and quantity of products produced. In order to deal with this problem, a proper classification and identification of leaf disease is required. The vegetable leaves might be affected by various diseases such as viral, fungal and bacterial. When the infection occurs on the leaves of the plant, the symptoms can be understood through change in texture, colour, shape and size of plant leaf which can further help in disease detection. Most of the symptoms are microscopic, so the identification of diseases is not possible due to the limited capabilities of human vision. However, it is necessary to develop an efficient technique which can help in detecting disease symptoms using scientific knowledge and experience. For this paper, the plant leaf images of tomato and potato are collected from Kaggle datasets. Image pre-processing and image processing steps are applied on the collected tomato and potato leaves. Along with these, various techniques such as image acquisition, image restoring, image segmentation, image augmentation, feature extraction and classification are performed for detection of plant diseases. In the pre-processing phase, the colour conversion technique is applied on RGB images which are converted into grey images. Several contrast enhancement algorithms are used to increase the quality of images after removing different types of noise from them. Flipping, cropping and rotating ways of image augmentation techniques are used and various properties such as portion, colour information or boundaries are traced in the image. In this paper, convolutional neural network (CNN) models such as AlexNet and ResNet-50 are used. AlexNet and ResNet-50 help in classifying the healthy and unhealthy leaf images by recognizing various diseases of leaves. According to the proposed framework, the image processing technique is performed on leaf datasets for disease detection. Next, the processed leaf images are classified using AlexNet and ResNet-50 architectures and finally, the overall leaf disease classification accuracy is analysed.

B. Paper II - Plant Leaf Detection and Disease Recognition using Deep Learning

A CNN - convolutional neural network is a deep learning model that is used prominently in image processing. The main idea of this paper involves three steps: First: acquisition of data. Second- pre-processing of data Third: image classification. The paper utilized dataset from Plant village dataset that contains plant varieties of sugarcane, apple, grapes, corn, potato, and tomato. There are 11 kinds of plant diseases detected in the study which also includes healthy images of identified plants. Image pre-processing involves re-sizing of images and enhancement before passing it for the classification model. The third step classification involves the main process which is convolution and pooling layers. The classification process classifies the plant leaf if it is infected with the disease or not, identifies the type of plant disease and recognize the plant variety. We have to test random images of plant varieties and diseases and achieve a considerable accuracy rate using many epochs while training the model.

III. METHODOLOGY

A. Paper I - Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures

Image processing techniques such as image pre-processing, image augmentation, feature extraction, feature selection and classification are applied on leaf images of tomato and potato that are taken from Kaggle dataset. A supervised machine learning model has been deigned which trains the dataset image and extracts the data from it. A leaf disease detection system using two CNN architectures such as Residual Neural Network-50 (ResNet-50) and AlexNet has been introduced.

1) Leaf Image Dataset: The leaf images of tomato and potato are taken as samples from Kaggle dataset which contains healthy and unhealthy leaf images. The dataset contains more than 4000 specimens of leaf images that are affected by four types of diseases mainly. These diseases are classified as Potato early blight, Potato late blight, Tomato early blight, Tomato late blight. The dataset also includes 2000 sample images of healthy leaves to construct the leaf disease classification and detection model.

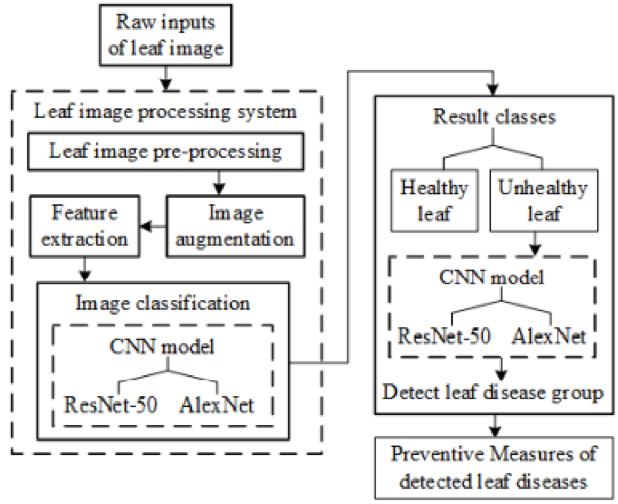

Fig. 1. Framework of Leaf disease detection with preventive measures

2) Image Pre-processing: The main objective of the image pre-processing technique is to transform the raw input leaf image into high quality leaf images and for eliminating the undesired portions from them. The various phases associated with this process are data cleaning, integration, reduction and transformation. The data cleaning phase deals with the elimination of the undesired distortion and rectification of inconsistent data. At the integration stage, heterogeneous data as well as data redundancy in leaf image datasets leads to data retrieval strategies which resolve multiple data conflicts and arrange a unified representation of data. Data reduction process helps in decreasing a large volume of data to increase the performances and efficiency of image processing. Data transformation involves data smoothing, aggregation, feature construction, data normalization and discretization. These leaf images are resized and converted into a dimension of 256×256 for training and testing datasets.

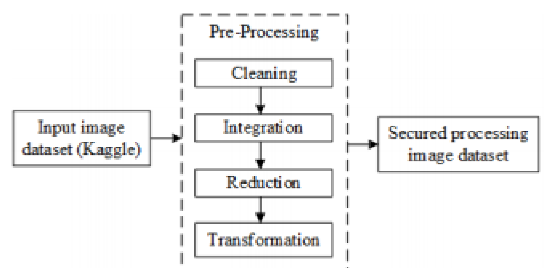

Fig. 2. Phases of leaf image pre-processing

3) Image Augmentation: Image augmentation is carried out for changing and modifying the representation of leaf images to accurately identify the diseases. To eliminate the chance of over-fitting and to enrich the simplification of the model, the training and testing leaf image datasets are augmented. The main role of image augmentation process is to resize the original leaf image dataset using techniques like flipping, cropping and rotation. It is also important for converting the leaf images into RGB using colour transformation technique.

4) Feature Extraction: The image feature vectors of the leaf disease can be extracted by the feature extractor of the CNN based detection framework. The properties of a leaf image such as colour, shape and texture are analysed through the feature extraction technique. The feature extraction technique helps to properly classify different leaf disease classes. For the leaf diseases, the feature extraction mechanism extracts the features of various lesion shapes and colours.

5) Classification based CNN Model: ResNet-50 and AlexNet architectures are used to identify the various diseases in the tomato and potato leaves. Classification-based CNN model in the image processing system with trained data and tested data of leaf images is used to categorize the leaf diseases class. AlexNet and ResNet-50 pre-trained network models are implemented to classify the leaf images of potato and tomato plants into various disease classes. 6000 different leaf images are classified into two different classes such as healthy and unhealthy leaf images through the AlexNet and ResNet50 architectures. 2000 healthy and 4000 unhealthy leaf images were obtained from the dataset. These architectures are further applied on unhealthy leaf images to categorize into four different disease classes such as potato early blight, potato late blight, tomato early blight and tomato late blight.

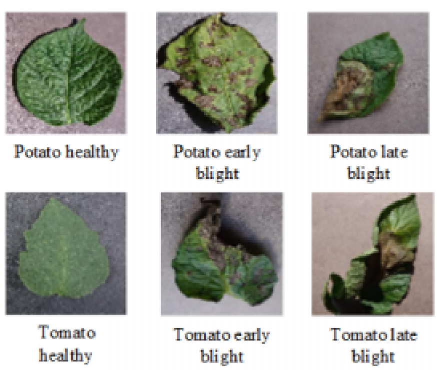

Fig. 3. Leaf diseases class of Potato and Tomato

B. Paper II - Plant Leaf Detection and Disease Recognition using Deep Learning

1) Convolutional Neural Network: Deep learning is a branch of Artificial Intelligence and machine learning that uses neural networks. Training the deep learning models involves dividing the feature extraction and extracts its features for classification. Some application of deep learning involves computer vision, image classification, restoration, speech, video analysis, etc. A CNN with basic process can simply detect and categorize. It is very efficient in evaluating graphical images and extracting the essential features through its multi-layered structure. CNN involves four layers, they are:

- Input image
- Convolutional layer and pooling layer
- Fully connected layers, and output.

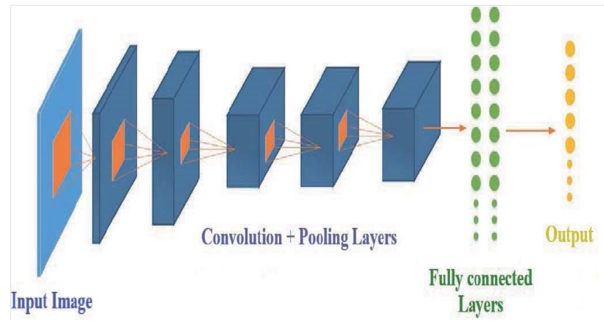

Fig 1. Illustration of CNN Architecture

2) Second Layer Convolutional Layer: Convolutional layers store the output of the kernels from the layer before which includes of weights and biases to be learned. The point of optimization function is the generated kernels that represent the data without an error. In CNN layer, a sequence of mathematical processes is implemented to extract the feature map of the input image. Fig. 2 shows the operation of the convolution layer for a 5x5 image input and a result is a 3x3 filter that reduced to a smaller size. Also, the figure also represents the shifting of filter starting from the upper left corner of the input image. The values for each step are then multiplied by the values of the filter and the added values are the result. A new matrix with the reduced size is formed from the input image.

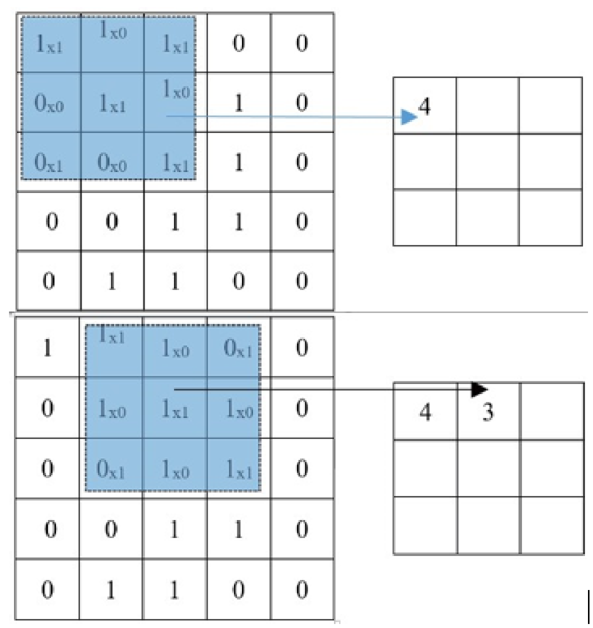

Fig. 2 A 5x5 input and 3x3 filter operation of convolution layer

3) Third Layer - Pooling Layer: This layer decreases overfitting and reduces the neuron size for the down sampling layer. Fig. 3 represents an example of the pooling operation. This layer reduces the feature map size, reduce parameter numbers, training-time, computation rate and controls overfitting. Overfitting is defined by a model by achieving 100 percent on the training dataset and 50 percent on test data. ReLU and max pooling were utilized to lower feature map dimensions.

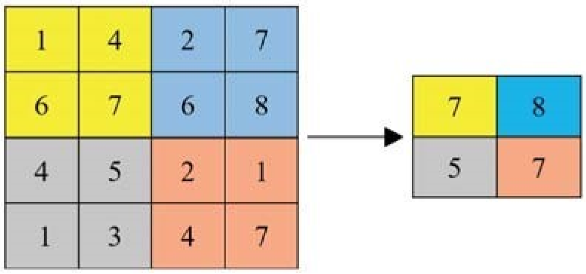

Fig. 3. Pooling operation

4) Fourth Layer - Activation Layer: Utilizes a non-linear ReLU (Rectified Linear Unit) activation layer in every convolution layer. application of dropout layers to prevent overfitting is also applied in this layer.

5) Final Layer- Fully Connected Layer: This layer is used to analyze the class probabilities and the output is the input of the classifier. Soft max classifier is the well-known input classifier and recognition and classification of sugarcane diseases are applied in this layer.

6) Steps: A block diagram presented in Figure shows the Input Dataset, Image Acquisition, Image pre-processing and Classification.

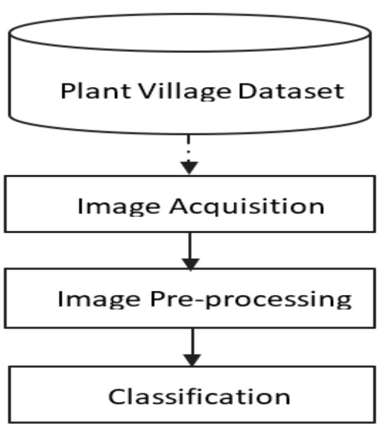

Fig. 4 Plant leaf detection and disease recognition methodology

- Image Acquisition: Dataset consisting of images used for training the model was acquired in the Plant Village repository. A python script was used to download images of the plant diseases from the repository. The acquired dataset consists of approximately 35,000 images with 32 different classes plant varieties and diseases.
- Image Pre-Processing: Images are pre-processed meaning reduced image size and image crop to a given input. It also adjusts the image to its needed color scale. The study uses colored and resized images to 96x96 resolution for processing.
- Classification: It uses a fully connected layers and for feature extraction it uses convolutional and pooling layers. The classification process classifies the plant leaf if it is infected with the disease or not, identifies the type of plant disease and recognize the plant variety.

7) Experimental Settings: Consisting of approximately 35,000 images containing 9 different types of tomato leaf diseases, 4 different types of grape leaf diseases, 4 different types of corn leaf diseases, 4 different types of apple leaf diseases, and 6 different types of sugarcane diseases. A neural network application program interface (API) written in Python was applied for the CNN model application. All the image dataset was used for training and testing uses 1,000 images that was taken from the field. Data augmentation techniques were integrated into the application to enhance the image dataset by rotating the images to 25 degrees, flipping and shifting of images horizontally and vertically. Adam optimizer is incorporated using a categorical cross-entropy. The model trained 75 epochs using a batch size of 32. All the experimentations were performed on Dell Inspiron 14-3476 i5 processor and memory size of 16GB.

IV. RESULTS AND ANALYSIS

A. Paper I - Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures

1) The analysis of healthy and unhealthy leaves using ResNet-50 architecture has been shown in the confusion matrix below. According to the results, the overall classification accuracy in this system for healthy and unhealthy leaf classes is 97.0 percent. The performance classification has been given in the table.

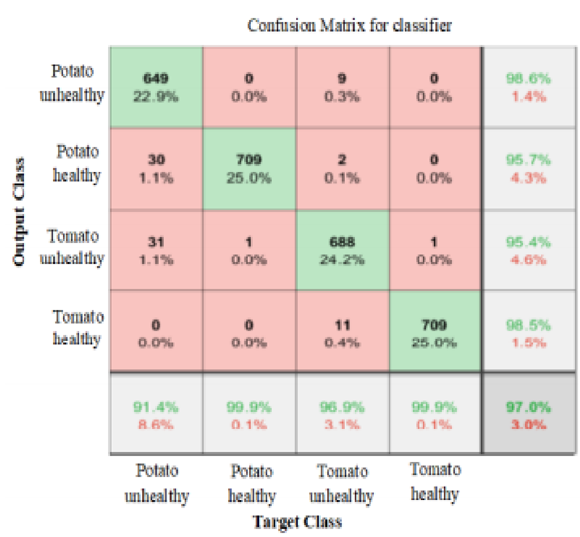

TABLE I
PERFORMANCE OF RESNET-50 MODEL FOR HEALTHY-UNHEALTHY LEAF

Classification label	Total no. of leaf image	Correctly classified images no.	Accuracy for Output class	Accuracy for Target class
Potato unhealthy	658	649	98.6%	91.4%
Potato healthy	741	709	95.7%	99.9%
Tomato unhealthy	721	688	95.4%	96.9%
Tomato healthy	720	709	98.5%	99.9%

2) The diagnosis of leaf diseases from unhealthy leaves using ResNet-50 architecture has been shown in the confusion matrix below. The results of this system show that the overall classification accuracy for the leaf disease classes is 96.1 percent. The performance classification has been given in the table.

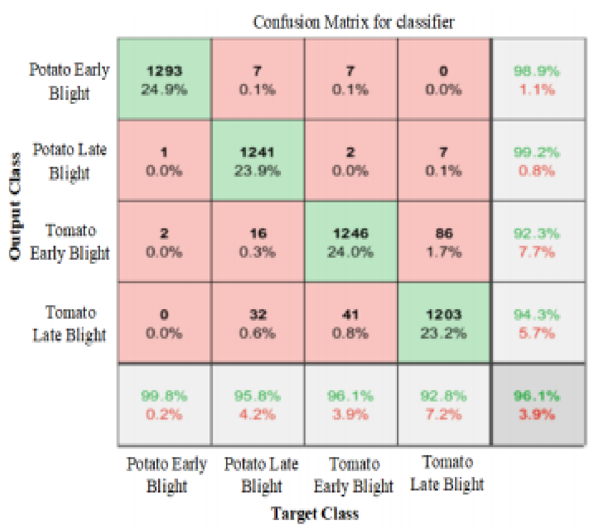

TABLE II
PERFORMANCE OF RESNET-50 MODEL FOR LEAF DISEASES

Classification label	Total no. of leaf image	Correctly classified images no.	Accuracy for Output class	Accuracy for Target class
Potato early blight	1307	1293	98.9%	99.8%
Potato late blight	1251	1241	99.2%	95.8%
Tomato early blight	1350	1246	92.3%	96.1%
Tomato late blight	1276	1203	94.3%	92.8%

3) The analysis of healthy and unhealthy leaves of potato and tomato plants using AlexNet model has been shown in the confusion matrix below. It is evident from the results that the overall classification accuracy for healthy and unhealthy leaf classes is 96.5 percent for this system. The performance classification has been given in the table.

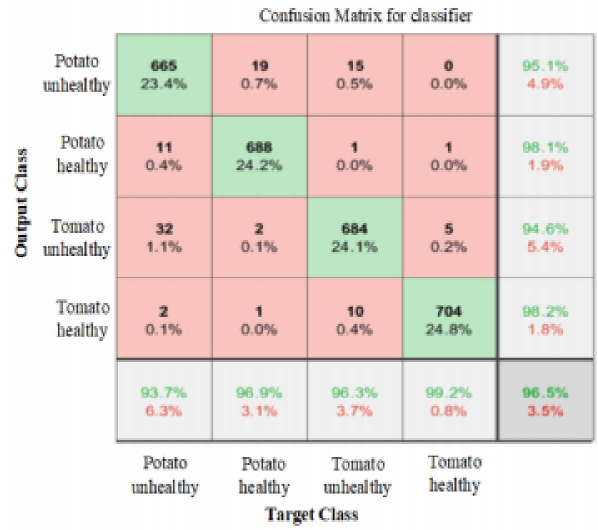

TABLE III
PERFORMANCE OF ALEXNET MODEL FOR HEALTHY-UNHEALTHY LEAF

Classification label	Total no. of leaf image	Correctly classified images no.	Accuracy for Output class	Accuracy for Target class
Potato unhealthy	658	665	95.1%	93.7%
Potato healthy	741	688	98.1%	96.9%
Tomato unhealthy	721	684	94.6%	96.3%
Tomato healthy	720	704	98.2%	99.2%

4) The diagnosis of potato and tomato leaf diseases from unhealthy leaves using AlexNet model have been shown in the confusion matrix below. It is observed from the results that the overall recognition accuracy for the leaf disease is 95.3 percent. The performance classification has been given in the table.

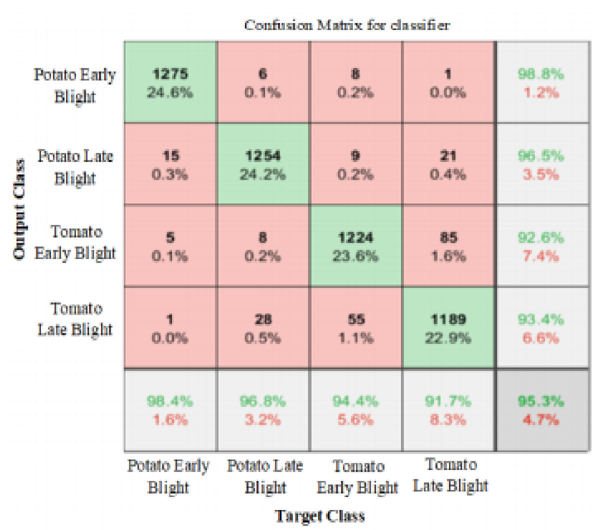

TABLE IV
PERFORMANCE OF ALEXNET MODEL FOR LEAF DISEASES

Classification label	Total no. of leaf image	Correctly classified images no.	Accuracy for Output class	Accuracy for Target class
Potato early blight	1290	1275	98.8%	98.4%
Potato late blight	1299	1254	96.5%	96.8%
Tomato early blight	1322	1224	92.6%	94.4%
Tomato late blight	1273	1189	93.4%	91.7%

For the classification of healthy and unhealthy leaves, an overall accuracy of 97 percent is achieved through ResNet-50 whereas an overall accuracy of 96.5 percent is achieved through AlexNet. For the leaf disease detection, the overall accuracy through ResNet-50 is 96.1 percent and the overall accuracy through AlexNet is 95.3 percent.

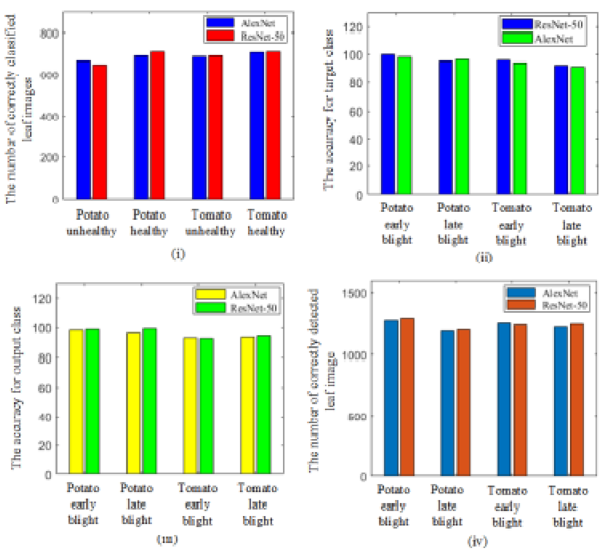

Fig. Comparison between AlexNet and ResNet-50

B. Paper II - Plant Leaf Detection and Disease Recognition using Deep Learning

A 96.5 percent accuracy rate was achieved using 75 epochs during the training of the model. The model also achieved a maximum accuracy rate of 100 percent when testing random images of plant varieties and diseases. The visualization of plots of train and test accuracy is described in fig 5. shows the model is effective in detecting and recognizing plant diseases. Fig. 6 shows of detection and recognition of a corn plant with 100 percent accuracy and it shows an accuracy rate of 100 percent recognition of healthy plant leaf on the left image and 99.56 percent affected with gray leaf spot disease on the right image. Fig. 7 shows the result of 99.77 percent and 99.58 percent accuracy of detecting and recognizing a tomato plant and it shows a 99.62 percent accuracy rate that it is infected with a late blight disease on the left image and 75.36 percent infected of early blight disease on the right image. Fig. 8 shows a 100 percent result of detection and recognition of a grape plant and shows a 100 percent rate that the leaf is infected with a late blight disease on the left image and 95.09 percent infected with a black rot disease on the right image. Fig. 9 shows the result of detection and recognition of an apple plant with 100 percent accuracy and shows a 100 percent result that the leaf is infected with a black rot disease on the left image and a 99.99 percent that it is a healthy leaf on the right image.

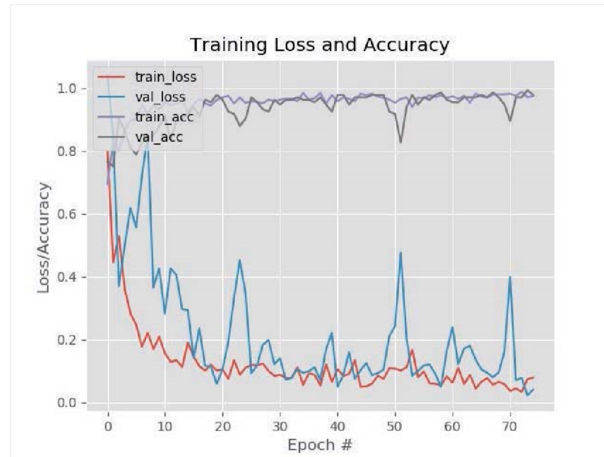

Fig. 5 Accuracy and loss against epochs

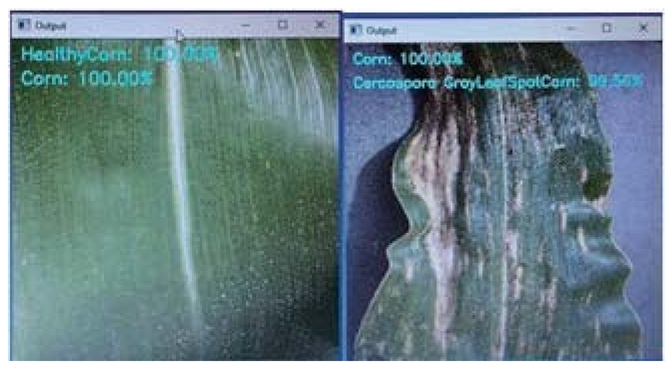

Fig. 6 Result of detection and recognition of a corn plant with 100 percent accuracy and shows a healthy plant leaf on the left image and diseased infected plant on the right image

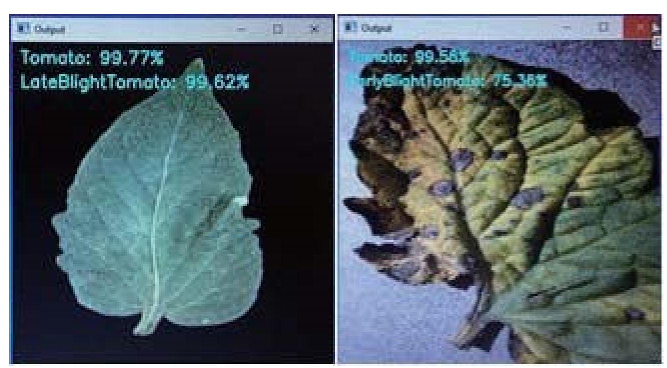

Fig. 7 Result of detection and recognition of a tomato plant with 99 percent accuracy and shows a leaf infected with a late blight disease on the left image and early blight disease on the right image

Fig. 8 Result of detection and recognition of a grape plant with 100 percent accuracy and shows a leaf infected with a late blight disease on the left image and black rot disease on the right image

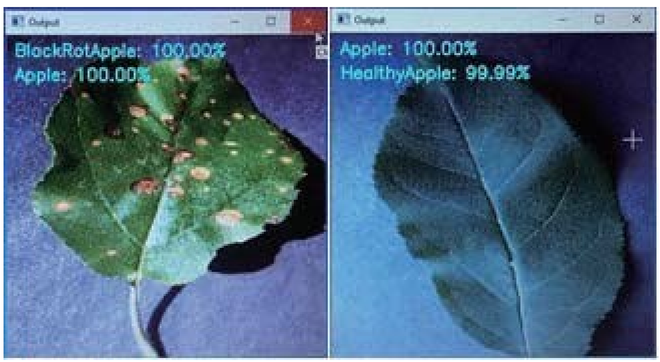

Fig. 9 Result of detection and recognition of an apple plant with 100 percent accuracy and shows a leaf infected with a black rot disease on the left image and a healthy leaf on the right image

V. Conclusion

A. Paper I - Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures

Leaf diseases are a major problem for farmers in the agricultural sector. Its timely detection is crucial to prevent its growth. A significant diagnostic approach using image processing and CNN has been suggested in this paper. It proposes an effective method which involves image processing techniques that are performed on images of Kaggle dataset for detecting and investigating the symptoms of unhealthy leaves. Moreover, this framework classifies the processed leaf images into potato early blight, potato late blight, tomato early blight and tomato late blight using AlexNet and ResNet-50 architectures. The overall classification accuracy of leaf diseases is analysed in the paper. It is hence concluded that the ResNet-50 model which has an accuracy of 97 percent is better than the AlexNet model that has an accuracy of 96.5 percent.

B. Paper II - Plant Leaf Detection and Disease Recognition using Deep Learning

Crops are the basic need for food and the people around the world rely on the agricultural industry immensely. Agricultural industry will benefit a lot from this early recognition and detection of these diseases. Using CNN, this study has achieved its goal to detect and recognize 32 different plant varieties and plant diseases. With the help of trained model, we can test real-time images to detect and recognize plant diseases.

VI. Future Scope

A. Paper I - Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures

Agriculture is one of the most important assets for each and every country. The growth of an economy majorly depends on the development in the field of agriculture. But there are different factors that can lead to the deterioration of the harvest and affect the quality and quantity of grains and vegetables. These grains and vegetables often come in contact with different diseases due to changes in climate and conditions in different places. As a result, cultivators can face severe losses because of these diseases. In order to deal with this issue, a robot which can be controlled using an android application can be created. This robot can be integrated with a machine learning model for autonomous driving. A camera can be fixed which can be used for navigational purpose. The robot can initially be trained to navigate through the entire field and later it can be allowed to move independently and take appropriate actions. The images of crops can be captured using the camera provided on the robot. To acquire high quality results, these images can be further processed using various techniques like image pre-processing, processing, segmentation and feature extraction. The processed images can be further passed through the models that have been implemented in the paper. This will help in detecting diseases in various plant leaves. The robot can further alert the farmer and appropriate measures can be taken on time to mitigate its spread.

B. Paper II - Plant Leaf Detection and Disease Recognition using Deep Learning

For scope in future is adding additional plant varieties and different types of plant diseases in the existing dataset to increase the trained models. Other CNN architectures may also use different learning rates and optimizers for experimenting the performance and accuracy of the model. With the achieved accuracy of 96.5 percent, the proposed model can assist farmers to detect and recognize plant diseases. This accuracy percentage can also be optimized.

References

[1] Husnul Ajra, Mst. Khairun Nahar, Lipika Sarkar, Md. Shohidul Islam "Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures", 21-22 Dec. 2020, 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE), IEEE, DOI: 10.1109/ETCCE51779.2020.9350890

[2] Sammy V. Militante, Bobby D. Gerardo, Nanette V. Dionisio, "Plant Leaf Detection and Disease Recognition using Deep Learning", 3-6 Oct. 2019, 2019 IEEE Eurasia Conference on IOT, Communication and Engineering, IEEE, DOI: 10.1109/ECICE47484.2019.8942686