Plant Leaf Disease Recognition Using Convolutional Neural Network

BY

Jannatul Mawa Hena
ID: 181-15-10938

AND

Novana Nasir
ID: 181-15-10865

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Md. Zahid Hasan
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2021

APPROVAL

This Project/internship titled "**Plant Leaf Disease Recognition Using Convolutional Neural Network**", submitted by **Jannatul Mawa Hena**, ID No: 181-15-10938 and **Novana Nasir** ID No: 181-15-10865 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 06-12-2022

BOARD OF EXAMINERS

______________________ **Chairman**

Dr. Touhid Bhuiyan

Professor and Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Internal Examiner

____

Dr. Fizar Ahmed

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Nusrat Jahan

Internal Examiner

Nusrat Jahan

Senior Lecturer

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

External Examiner

Dr. Mohammad Shorif Uddin

Professor

Department of Computer Science and Engineering

Jahangirnagar University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md. Zahid Hasan, Assistant professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Md. Zahid Hasan
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Jannatul Mawa.

Jannatul Mawa Hena
ID: - 181-15-10938
Department of CSE
Daffodil International University

Novana

Novana Nasir
ID: - 181-15-10865
Department of CSE

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible to complete the final year project/internship successfully.

We are really grateful and wish our profound indebtedness to **Md. Zahid Hasan**, **Assistant professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "**CNN**" to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Professor Dr. Touhid Bhuiyan, Professor, and Head** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Plants are really important for the planet and for all living things and the leaf produces food for the plant by photosynthesis. Also our country, Bangladesh, is an agriculture dependent country. So to save our plants from diseases, we need to detect the plant leaf diseases too. Food production decreases because of plant diseases. So, there is increasing scarcity of foods. For this reason, detecting those plant diseases is important. To classify the disease with only eye observation cannot be correct. Sometimes laboratory tests cannot bring exact results. Nowadays Deep learning methods play an important role in detecting objects. Besides, the Convolutional neural network (CNN) has taken a good area to detect plant diseases. We have solved the problem by using a CNN method (resnet34) and using some layers to predict the diseases. Here, we have summarized DL laws to get accurate predictions of diseases. We also have shown here some CNN principles (resnet34). Progressing the idea of plant disease recognition, will greatly help our farming area. On the other hand, nutrition problems can also be solved and increase the crops cultivation. We have also discussed the future plan and future work about plant diseases recognition.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

In the agricultural sector, automatic plant leaf disease identification plays a vital role. [1]. Nowadays higher inventions have enabled people to supply the excellent amount of nutrition and food required to fulfill the requirements of the growing population. If we talk about Bangladesh, we know that our country is mainly an agricultural country. About 85% of the Bangladeshi people are directly or by suggestion dependent on agriculture as the soil of Bangladesh is fertile and also the climate is favorable for cultivating. Fruits and vegetables are the common items and also the fundamental agricultural things. So, in the field of agriculture, detection of disease in plants plays an important area. For the reasons of cultivating large numbers of crops, even the agriculturists sometimes fail to detect leaf diseases by observing in naked eyes on the disease-affected leaves [1]. Although, within the countryside of developing countries, eye observation remains the first approach of disease identification [2]. Rural areas people are not aware about diseases properly .Sometimes they try to find the solutions. So they have gone to the consultant too. To overcome the above problems, various types of solution sets can be used in deep learning and machine learning for the classification of plant diseases. Also we need to know the types of plant diseases in order to train our machine to identify the type of disease it has, also if the leaf is healthy or not. In the recent Years, Deep Learning has led to great inventions in various fields like Image Recognition, image processing, Speech Recognition, and Natural Language Processing and many more.

Using the Convolutional Neural Network in the problem of Plant leaf Disease Detection has very good results. CNN is known as the best recognition method for Object Recognition.

The goal of this study is to classify plant leaf disease using a Deep learning approach along with computer vision. Surroundings we can see mainly Powdery and Rust type diseases. Rusts are a type of plant disease, with a particular focus on their leaves caused by pathogenic fungi attacks. At present, there are approximately 168 genera of rust and 7,000 species and belong to the genus Puccini(over a half) [3]. Plants with serious rust infection may appear small, chlorotic (yellowed) or may display signs of infection such as rust fruiting bodies. On the other hand, powdery disease refers to the fungal attack on a wide range of the plant's leaf. The order of Erysiphales is caused by this disease. Powdery mildew is one of the commoner plant diseases to identify. Its symptoms are quite different. The infected area displays white powdery spots on the leaves. This also causes harm for plants. Besides, it is again important to be conscious of powdery diseases. So we can find that it is definitely very important to identify rusts, powdery and healthy leaves of the plant's leaves.

Deep learning is a subset of machine learning in AI that intimates the human brain to behave similarly while processing data and producing patterns. This study's goal is to classify the plant disease type and also whether the plant is healthy or not using the Convolutional Neural Network. CNN is known to be the most powerful visual model in computer vision for empowering accurate segmentation by yielding hierarchies of features. Due to the ongoing disclosure of better results, the multiple layers supervised network has become favorable among researchers [4]. They are working very fast as the previous algorithms for good working performance simultaneously.

1.2 Problem Statement

A plant leaf disease is an area of function that interrupts its vital works. Plants of all kinds, wild and cultivated, are susceptible to disease. Although each species is susceptible to specific diseases, there are only a few of them in each case. Plant diseases vary seasonally, unexpected occurrences also the environment conditions. Some diseases happen with the outbreak of plants. Some plant varieties are more susceptible to disease outbreaks, while others are more resistant. Image processing has had a vital impact on our technological progress. However, in order to use it properly, we must first understand why, where, and how to use it. If we want to find a solution, we must first identify the problem. We need to learn everything we can about the problem if we want our solution to be as effective as possible. We must also determine the work's requirements in order to accept it as a solution besides, To identify easily the disease of plants and find the solutions at an early age.

1.3 Research objective

The objectives of our work are given below:

1. To construct a model of existing machine vision-based systems that can correctly classify different types of plant leaves into their respective classes.
2. Increasing accuracy of detection of plant diseases.
3. Increase the use of technology in agriculture.
4. Reduce the cost of cultivation.
5. Help farmers identify diseases and take precautions as early as possible.

1.4 Research Questions

The main questions on which this thesis is focused are given below:

1. The state of image processing implementation in agriculture at the moment.
2. The limits of image processing in the identification of leaf diseases.
3. The procedure for resolving these issues.
4. How does CNN use it to find accuracy?
5. How does it classify the class of diseases of plants?

1.5 Research Layout

Chapter 1: will cover the following topics: introduction, problem statement, research objectives, and key research questions.

Chapter 2: will highlight the detailed review of related works as well as the current state from the perspective of Bangladesh.

Chapter 3: the approach of the suggested model in the sector of agriculture in Bangladesh will be depicted with a detailed description.

Chapter 4: presenting the result analysis and comparison of existing works.

Chapter 5: describes the conclusion of this research along with shows a path for future work.

CHAPTER 2

Literature Review

2.1 Related Works

We can control our agricultural loss and at the same time increase the fertility by performing the appropriate techniques of classifying the healthy and unhealthy leaves. The Agriculture department has become a fundamental section of visual data classification of the development of our country. In the past, we have seen different researchers have done several approaches to identify plant leaf disease.

In paper [5], we see that the authors used tomato leaf images and applied several geometric and histogram based methods to the diseased part of the leaves. After that they performed an SVM classifier. The classifier contained several kernels to identify the disease.

The article [6] introduces a new model of plant leaves classification using CNN. They used two models here based on leaf discoloration and damaged leaf using Google Net. They got accuracy above 94%. If there is less than 30% damage to the leaf, their model can still detect it.

In this paper [7], Liu builds a model based on the Apple leaves images using a deep convolutional network that accurately identifies different types of properties within the Apple leaves. He used 13689 images and achieved a very high level of accuracy by applying a method of image processing, PCA oscillation. He also proposed a new AlexNet neural network, with the NAG algorithm that optimizes the network.

[8] Proposed a method that describes numerous ways for classifying infected leaves and plants. Here a model consisting of different levels, CNN is used. The whole process mentioned in this paper is based on leaf images and it is done using training, testing and image enhancement techniques (histogram alignment and contrast adjustment). They

transformed the RGB images to grayscale images and for that they used color conversion. The image enhancement techniques were used to improve the image quality. Classifications like SVM, ANN and FUZZY were also used here. Diseases of unwrapped tomato plants can be identified with ANN and FUZZY classification. Using these images we can accurately differentiate between diseases of different plants.

In [9], they proposed a model and their main goal was to reduce the damage caused by plant diseases in the early stages of crop production. They developed a model for the detection of tomato leaf diseases that tested six types of diseases, such as bacterial spots, late blight, leaf curl, mosaic virus and Septoria leaf spot. They used a dataset of 500 pictures. To remove background from the images they used image thresh holding which is a part of image segmentation. After that they run AND operations between an actual image and a mask image and get a segment image from it. They also performed feature extraction in two ways. They used two more algorithms, Naive Bayes Classifier & K Nearest Neighbor to detect tomato leaf diseases and get 65% accuracy from Naive Bayes Classifier and 80% accuracy from K Nearest Neighbor. From this it can be said that they get better results in KNN.

Besides distinguishing plant disease, Ferentinos et al. [10] acknowledged a method to detect 58 types of plant diseases which are different from each other. For this they used several CNN models and got an effective accuracy to identify plant diseases. They tested their CNN model with actual images.

In paper [11], Sladojevic builds a model to plant diseases. Their model was built to identify 13 different types of plant diseases. Here a DL architecture (Caffe DL framework) was used to do the process and also to execute the training of CNN.

In [12], Ramacharan described a method to identify 3 diseases and 2 pest-damage in cassava plants. Here they used a transfer-learning approach. Later they broaden their work and developed a CNN model based on smartphones and increased the rate of identification of diseases of cassava plants. They got 80.6% accuracy on the new model.

In the paper [14], the researchers suggested a CNN model to diagnose plant diseases. Their model has 9 layers and they used the Plant Village dataset for this experiment. Also they increased the number of images with the help of data augmentation approaches. The accuracy they got from this is better than traditional machine learning based approach.

In the paper [15], they discussed how to detect leaf disease of jackfruit plants from digital images. They mainly deal with 3 types of diseases which are- Rhizopus rot, pink disease and leaf spot. They used 480 images and resized them to a predefined format. Then Histogram equalization was used for image segmentation and contrast division of image. The K-means clustering method separates the diseased part of the leaf from the diseased-free part. Finally 9 high classifiers were used to diagnose the disease. Of these, Random Forest gives the highest accuracy which is 92.5% and KNN gives 70.42% accuracy which is the lowest compared to other classifiers.

From these discussions above, we can see that there are some similarities and dissimilarities of the above discussed works with our work. Our work is related to plant disease recognition based on leaf images using CNN and Deep Learning methods. A summary of the related work discussed above, is shown in Table 1.

Table 2.1 A summary of our related work with plant disease recognition based on plant leaves.

References	Methods	Results
Mohanty et al. [5]	GoogleNet and AlexNet	99.34% accuracy in GoogleNet 99.27% accuracy in AlexNet
Jeon [6]	CNN and GoogleNet	94% accuracy
Liu, Bin et al. [7]	XDNet and Xception	90.79% in XDNet 88.62% in Xception
K.Padmavathi and K.Thangadurai et al. [8]	Image pre-processing (RGB and Grayscale)	RGB image pre-processing gave better accuracy than Grayscale image pre-processing
Subham Malvankar et al. [9]	Naive Bayes Classifier and KNN	65% accuracy in Naive Bayes Classifier 80% Accuracy in KNN
Ferentinos [10]	CNN (AlexNetOWTBn and VGG)	Accuracy in AlexNetOWTBn is 99.49% Accuracy in VGG is 99.55%
Sladojevic [11]	CNN (Finetuned)	96.3% of accuracy
Ramcharan [12] (2017)	GoogleNet based Inception V3	accuracy of 93%
Ramcharan [13] (2019)	SSD model with MobileNet detector and classifier	accuracy on images 80.6% accuracy on video 80.6%
Geetharamani [14]	9 layered deep CNN	96.46% on accuracy
Md. Tarek Habib, Md. Jueal Mia et al. [15]	Random forest and KNN	92.5% accuracy in Random forest 70.42% accuracy in KNN

2.2 Bangladesh Perspective

From Bangladesh's viewpoint, a high range of population has been aware of raising crops. Our administration is trying to guide our producer but they do not have a much understanding of technology. The economy of our country is fully dependent on agriculture and also on the growth of the harvest every year. So, plant disease recognition has become the most critical thing for cultivating crops. If the identification is not correct or late, it will cause uncontrolled loss to the growth of crops. It also has an impact on the financial status of the farmers. The govt. does not offer enough force into examining our farming difficulties, significantly plant disease detection obstacles. So, the people are not conscious of it. The management does not execute appropriate workshops to provide enough common awareness about leaf disease issues. Due to lack of adequate knowledge about technology, producers have no interest in technology. There is a different dilemma that they are not familiar with technology. As Bangladesh is an agricultural country, we must have scientific technologies and our farmers should have a proper knowledge of this serious problem.

2.3 Scope of the Problem

Problems occur when our farmers are not understanding the technologies. Sometimes it shows that only the eye's viewpoint can't be corrected. The government does not devote enough resources to investigating our farming problems, notably the challenges of detecting leaf disease. As a result, no one will be aware of it. The management does not conduct appropriate workshops to raise public knowledge about leaf disease concerns. Producers have little interest in technology due to a lack of appropriate information about it. They have a distinct problem in that they are not used to using technology.

2.4 Challenges

a) **Data collections**: It was not easy to find rusty, powdery leaves easily. Sometimes the collections of rust data increase, sometimes the powdery data increase. Therefore, it was very hard work to collect the images from the plant leaf's field.

b) **Raw data processing**: While collecting the images from various sources, sometimes images are noisy, sometimes it's tough to process these images perfectly and then divide these images in 3 classes.

c) **Selecting machine learning approach**: Several types of machine learning approach are there. Many researchers used different types of algorithms . So selecting the perfect algorithm is more challenging as well.

d) **Accuracy improvement** : Researchers make their result by their model. In the same topic, there is much research. So, improving the accuracy by work is kind of another issue.

CHAPTER 3

Methodology

3.1 Working Process

In this part, we will discuss the whole process we build to recognize plant disease based on plant leaves using CNN. The working process of our work in this project is shown in the following steps of the figure (figure1) below:

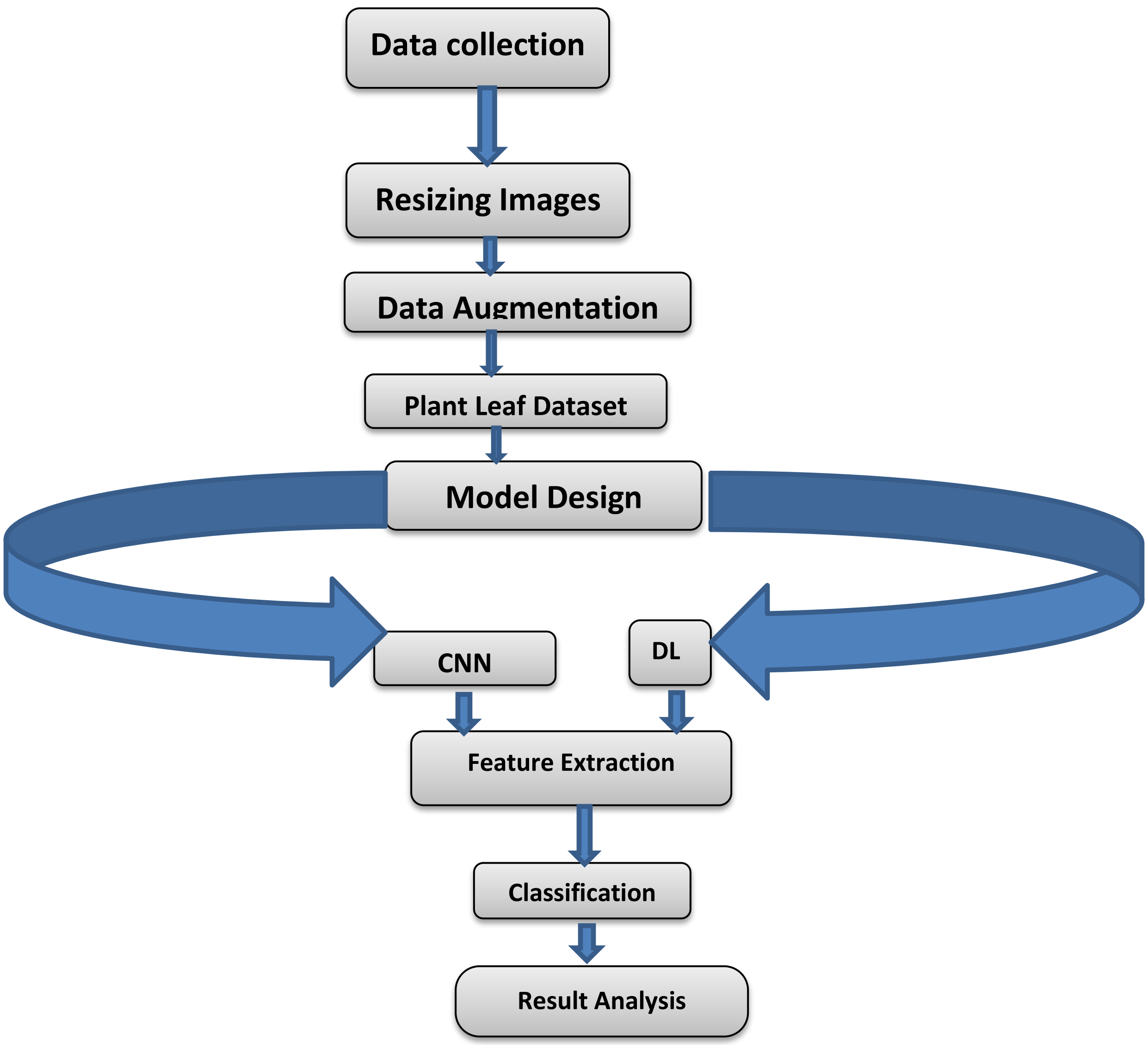


Figure 3.1 Working process of the proposed methodology

3.2 Dataset Presentation

In this study, we used a dataset of 1532 images and we had collected all primary data by digital camera. [16] Also take some more pictures from website, so that we can get more accuracy and train more data as well. These images from a public plant leaf datasets from the website [17]. In this dataset there are images of two types of the disease affected plant leaves (rust & powdery) and also the healthy leaves. All the images are divided into three parts. Those are training, test and validation sets. Images are then kept in Google drive.

Table 3.1 Individual part of each class:

Class	Healthy	Powdery	Rust
No. of test images	**50**	**50**	**50**
No. of train images	**458**	**430**	**434**
No. of validation images	**20**	**20**	**20**

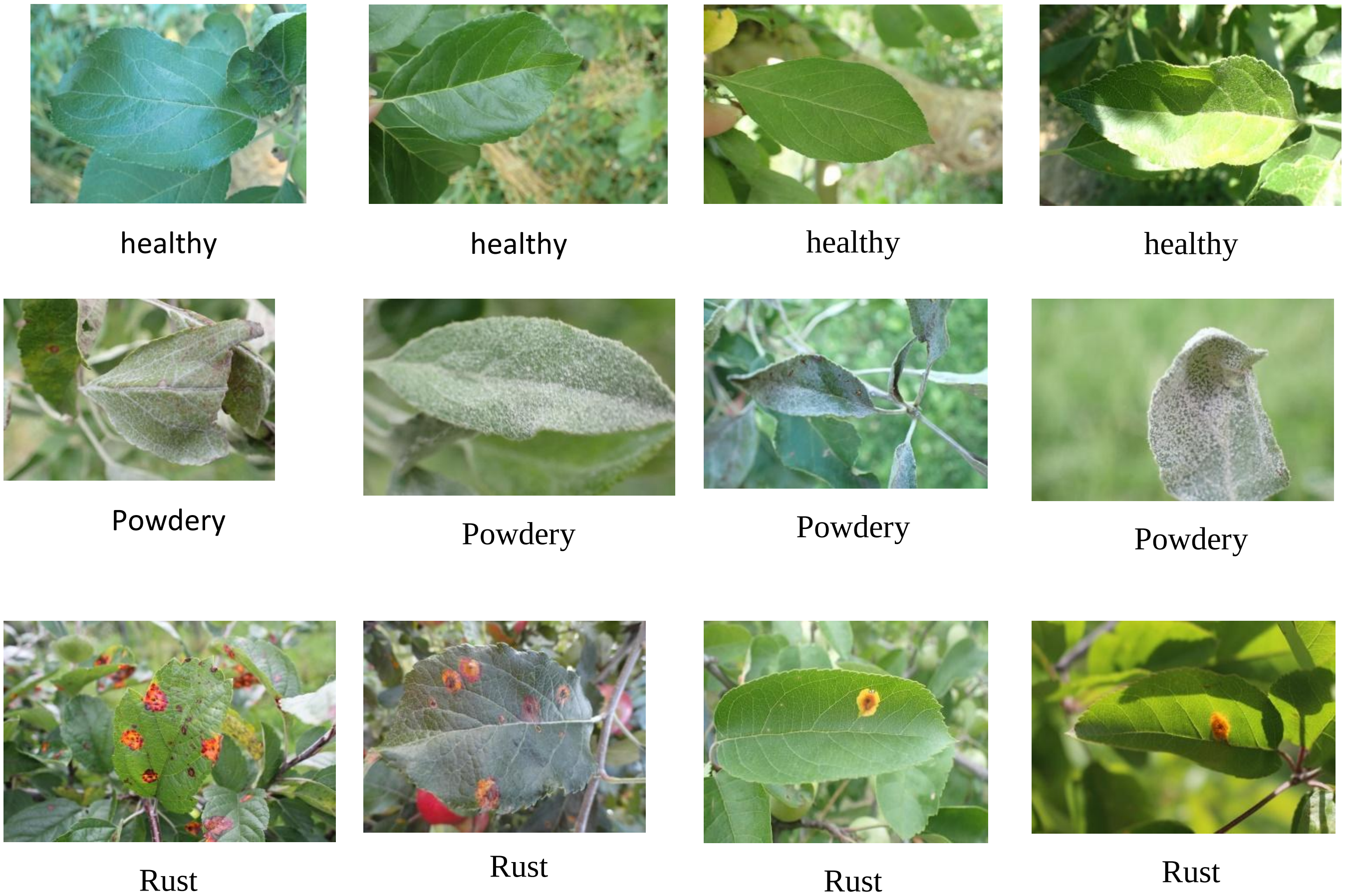
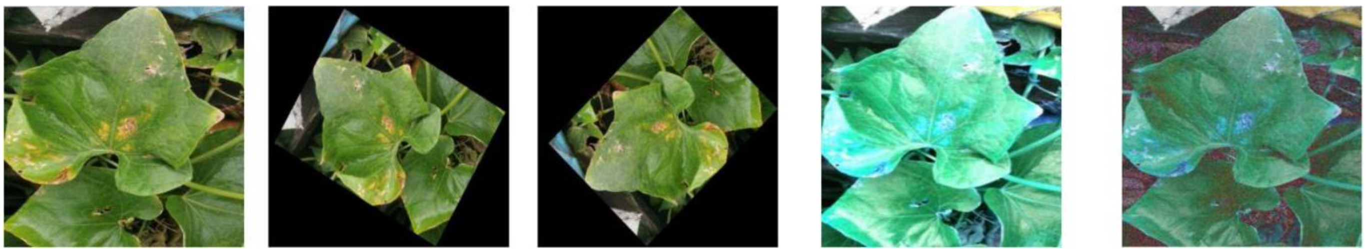


Figure 3.2 Healthy, Powdery and Rust leaves

All of the leaves were collected in a pleasant atmosphere from various agricultural sites located across Dhaka. Those collections of images are not properly suitable cause here some of them are noisy and color fluctuation. To overcome those problems, data augmentation, rotation, scaling, transformation e.t.c. methods were applied.

3.2.1 Data Augmentation

In this study, we used a dataset of 1532 images. And we had augmented these images. After augmentation Image increases twice. In this dataset there are images of two types of the disease affected plant leaves (rust & powdery) and also the healthy leaves. All the images are divided into three parts. Those are training, test and validation sets. Images are then kept in Google drive.



a) Original **b**) Rotate 45° **c**) Rotate 135° **d**)Adjust brightness **e**) Add gaussian noise

f) Mirror linearization **g**) Texture enhancement **h**) Adjust saturation **i**) Adjust sharpness **j**)Histogram

Figure 3.3 Traditional image data augmentation methods.

Among all the benefits of CNN the main benefit is their data processing ability. CNNs can process data which has never been seen before. But when the size of the data is not sufficient, they choose to over-fit the training data, this means the data generalization is not possible [23]. So, Over fitting is reduced with data augmentation technology and the dataset is enlarged to reproduce the image's (plant leaf), light, exposure, noise and angle during the image preprocessing step. Image augmentation method is performed on the training dataset only. We have seen that Liu [7] used the image augmentation methods to increase the apple diseased leaf images.

The images are processed by increasing and decreasing the brightness, contrast and sharpness by a certain range. Here we also used rotation methods by 45, 90°, 180°, 135°. And then we have done mirroring, flipping (both horizontally and vertically) and symmetry operations too. Also noise and blur elimination is also very important in plant leaf images so that the model can detect the diseases correctly. A Gaussian filter is executed for this purpose as,

$$\boxed{G (x, y) = x2+y2\ e2\ 1/2\pi\ (\sigma 2\ )\ (\ e - (x2+y2/2e2))}$$

The rotation of an image is done by pixel rotation. It is done at an equal angle depending on the center. Suppose, the center point is C (a, b) and after performing rotation to ϴ° the new point is C2 (a, b). Then the two points are calculated as,

$$\begin{gathered} a = r \cos \alpha \\ b = r \sin \alpha \\ X = r \cos (\alpha - \Theta) = a \cos \Theta + a\ r \sin \Theta \\ Y = r \sin (\alpha - \Theta) = -a \cos \Theta + a\ r \cos \Theta \end{gathered}$$

After performing data augmentation, the size of our training dataset has increased to 2000 images.

3.3 Image Pre-processing

Image pre-processing steps are used to format images. These steps are performed before they are used by the model's training and inference. By this technique we can transform raw data into an understandable format. Raw data means the real world data. Raw data is never completed and we cannot send incomplete data through a model because it will refer to certain errors. So, this is the reason we have to pre-process data before sending that through a model. Image pre-processing includes resizing, orienting and color corrections. These techniques are applied to training and testing datasets. The original data which is seized through the sensor containing a predominance image typically represented via an image matrix function values (brightness) is impersonated by the output images. A given dataset may contain images that are generally less contrasting. Each image must go through a fixed amount of contrast adjustment to improve the model performance, if the model is used in all situations to produce only low contrast images. Even though mathematical changes of images (rotation, scaling, and translation) are sorted among pre-processing strategies; comparative methods are utilized here too.

3.3.1 Image Enhancement Technique

The technique of improving the quality of an image which is recognized by a human is known as Image enhancement techniques. There are a lot of techniques for improving the quality of images. These are applied one by one to each of the bonds of a multispectral image. In an image, the scale of brightness values can be increased by contrast enhancement techniques as the image can be fluently presented in an expected manner. There are two types of contrast enhancement techniques which are linear and non linear transformation.

The histogram equalization technique is known as grayscale transformation for contrast enhancement. It is a non-linear contrast enhancement technique and it can improve contrast in images and also represents the number for each strong quality pixel. But this technique can't be applied individually to the RGB parts of the image cause it will result in a drastic change within the image’s color balance. Histogram adjustment may create an outcome that is more awful than the first picture as the histogram of the subsequent pictures turns out to be roughly level But if first the image can be converted to another color space like HSL or HSV color space, then this algorithm can be applied without any change in the saturation and hue of the image. But many times it increases the contrast of background noise rather than decreases the usable signal.

Figure 3.4: enhancement technique in rust, powdery and healthy leaf

3.4 Convolution Neural Network

Deep learning (DL) approaches, specifically the ones which stand on convolutional neural networks (CNNs), are now widely used on agriculture for identifying and classifying plant diseases, weed identification[17], crop categorization and more. CNN may be a sort of artificial neural network that has one or a lot of convolutional layers and square measure used for image process, recognition, classification, etc. and conjointly for alternative co-correlative knowledge. As computers see an image as a matrix of numbers that represent a single component, it's essential that the relation between the pixels remains the same after the image is handled through the network. CNNs square measure wants to save the contiguous relationship between the pixels that have totally different mathematical operations stacked on prime of every alternative to get the layers of the network. CNN is the most extensively used categorization system for plant leaf diseases.

3.4.1 Architecture of CNN model

In here, we've advised a three-track deep neural network model design. The first layer includes a deep learning algorithmic program, whereas the second layer is connected to each other. There are many more neurons to connect the layers. Besides, the ultimate targeted layer of output neurons, the neural network has hidden neurons too. That is four layers. The input holds 160×160×4 neurons, which represents the RGB(Red, inexperienced and Blue) value for a 160×160×4 image. the primary convolution-pooling layer applies an area receptive field of size 4×4 with a stride length of one component to extract thirty two feature maps, followed by a scoop pooling operation managed in a very 2×2 region, the second and third convolution-pooling layers use an equivalent 4×4 native receptive field (kernel) ensuing sixty four and 128 features maps severally whereas alternative parameters stay fastened. The fourth layer is a fully-connected layer having 128 ReLU (rectifier linear unit) neurons. The three convolutional-pooling layers conjointly utilizes ReLU activation functions.

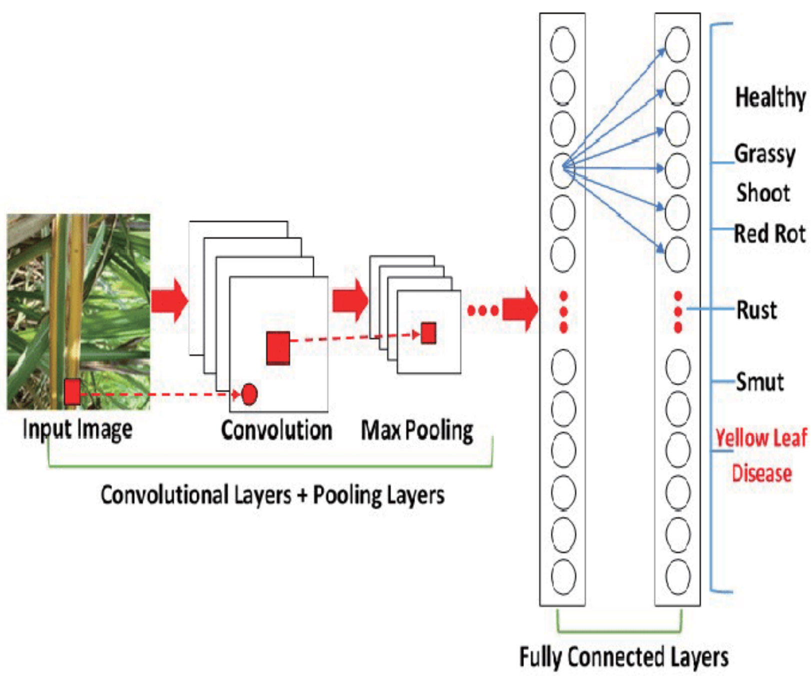
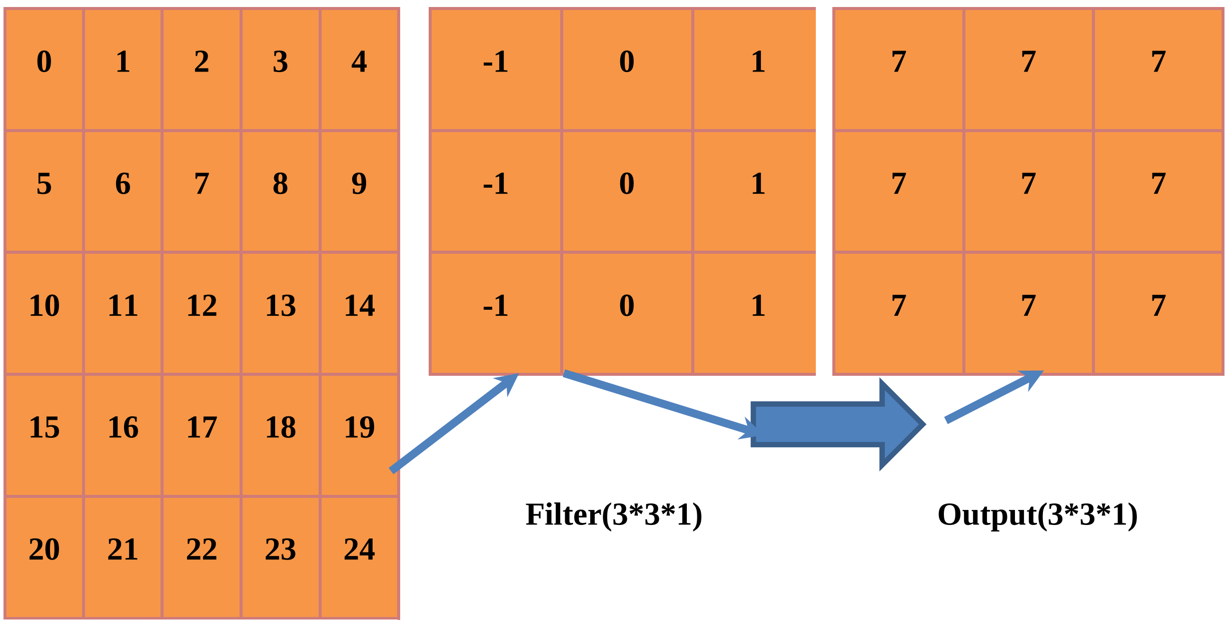


Figure 3.4 CNN for plant leaf disease recognition.

3.4.2 Convolution layer

Convolution layer is made by a number of filters and each number of filters have different types of parameters, those parameters need to be learnt. The filters' height and weight are less than that of the input volume.. It is the initial layer that takes an input picture and extracts characteristics from it. By learning visual attributes with tiny squares of input data, convolution retains the link between pixels. It's a mathematical process with two inputs: an image matrix and a filter or kernel.



Input(5*5*1)

Figure 3.5 The process convolution operation

3.4.3 Max pooling Layer

Max Pooling could be a convolution method wherever the Kernel extracts the maximum value of the area which it rotates. The purpose of pooling is to reduce maximum parameters of the filters. And that is making a set of different filters. So, it decreases the amount of parameters to learn and also the amount of computation performed in the network. As a result, following up these methods features map showing below:

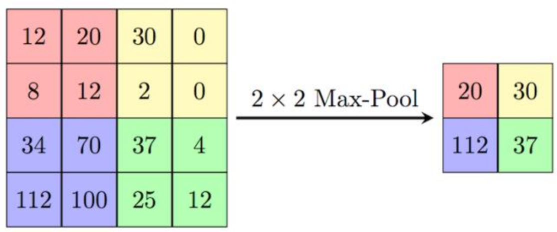


Figure 3.6 Max pooling layer

3.4.4 Fully connected layer

Fully connected neural networks are connected to each other. Its levels are the network's final layers. The weights matrix is a 9x4 matrix, while the input is a 1x9 vector. The output vector is gained by taking the dot product and applying the non-linear transformation using the activation function (1x4).

3.5 Artificial Neural Network

Artificial Neural Networks (ANNs) are mainly considered as computational dispensation structures that are highly enlivened by the way biological nervous structures (for example human brain) activate. ANNs are mostly included in a great number of consistent computational nodes (indicated to as neurons) that employ interweave in a distributed manner towards cooperatively gain from the input to improve its final output. ANN network is the heart of deep learning algorithms. The fundamental architecture of the ANN model is shown in the figure (num).

ANNs are most commonly used and there are three layers in them. The layers are input layer, hidden layer and output layers. Different types of layers work on different input layers. The signals travel from the primary layer (the input layer) to the last layer (the output layer), presumably when traversing the layers numerous times. We would weigh the input to the input layer generally as a multidimensional vector which will disperse it to the hidden layers. The hidden layers can at that time choose verdicts from the preceding layer and weigh up in what way an unpredictable variation inside itself obstacles or develops the final output which is implied to be the way towards learning. Absorbing numerous hidden layers weighted upon each other is often titled deep learning.

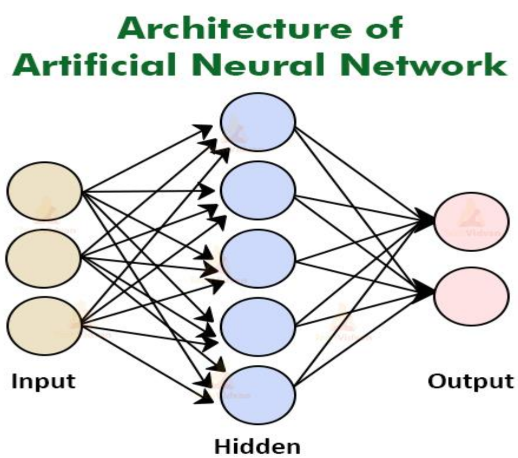


Figure 3.7 Architecture of ANN

3.6 Deep Learning

Deep learning (DL) is a technique that emulates the procedure human gain particular types of knowledge and a kind of machine learning as well as AI (artificial intelligence).Over the last several decades, Deep Learning has proved to be a strong tool to recognize images as it has the ability to manage giant quantities of data. People's interest in using the hidden layers has gone beyond traditional techniques, especially in pattern detection. Convolutional Neural Networks are considered to be one of the most famous DNN. We use DL especially to classify images, identify objects and natural language processing. It is an algorithm by which automatic characteristics of data choice can be done based on neural networks. It merges low-level characteristics to identify distributed properties and features of sample data and to create high-level characteristics for it. The accuracy and generalization of image classification in DL has been much better than the image classification done with the help of traditional methods. CNN is currently the most widely used DL method for diagnosing leaf disease. The rest of the deep learning networks are usually used for image segmentation [20, 21], medical treatment, etc. Diseases of plant leaves cannot be identified with these. The pixel values are coated with kernels and their equivalent kernel values are multiplied and then summated after that at the very end bias is added. Until all possible positions of the image are filtered, the kernel keeps moving by one pixel. Its convolutional layer and pooling layer usually appear preferably in application. Each neuron has a fully connected layer attached to the upper neuron [22]. Many CNN based classification models have been introduced (ResNet, GoogleNet, AlexNet, VGGNet, etc.) which have been created in deep learning evaluation.

3.7 Transfer Learning Approach

Transfer learning shows that previously trained datasets work on a new task by deep learning. It is useful in deep learning because a little amount of data trained by a network is already given. A little amount of data with a high degree of precision machine uses knowledge in transfer learning. To increase generalization about another task, use knowledge obtained from a prior activity. Dropout rates, learning rates, and batch sizes are all factors to consider. MobileNetV2's input size and MobileNetV2's output size EfficientNetB0 has a value of 224×224 [23]

3.7.1 ResNet(34)

It is called the backbone of vision tasks. ResNet was a game changer because it plays vital rules to train data effectively. It works by effectively training extraordinarily DNN (deep neural networks) with 150+ layers. Resnet34 is a state-of-the-art image categorization model with 34 layers of convolutional neural networks. This is a model that has been pre-trained on the Image Net dataset, which has 100,000+ photos divided into 200 categories. So it helps to detect photos easily.

3.8 Training and Testing

We have divided our training and validation dataset and our total images are around 1532. The whole dataset is divided into 80% for training and the remaining 20% for validation. Our model has been trained with the training set and the hyper parameters have been adjusted while training the model with the validation set. And some data is kept in the test set which was not used in training and validation. That is to say, this is a sample data which the model has never seen before and with this test set we can evaluate the efficiency of our model.

The model was trained with a transfer learning approach. Here categorical cross entropy is used as a loss function and it is shown in Equation (1). Our learning rate was set at 0.001.

$$i_{CE} = -\sum_{i=1}^{n} * t_1 log(p_1) \quad (1)$$

In equation (2), the activation function of the model is shown with Adam optimizer where SoftMax has been used.

$$\underline{f_1(a)} = \frac{e^{n1}}{\sum_k e^{nk}} \quad (2)$$

The complete working process of our work is shown in figure.1

The entire experiment is executed on two 64-bit Windows Operating Systems. One is the Intel Core i5-xxxxxxx and the other one is AMD E2-9000e RADEON R2, both CPU processors with 8 GB RAM and 4 GB RAM respectively and a 1 Terabyte hard disk using the python programming language in an anaconda environment.

Chapter 4

Experimental Result Analysis And Discussion

4.1 Result And Discussion

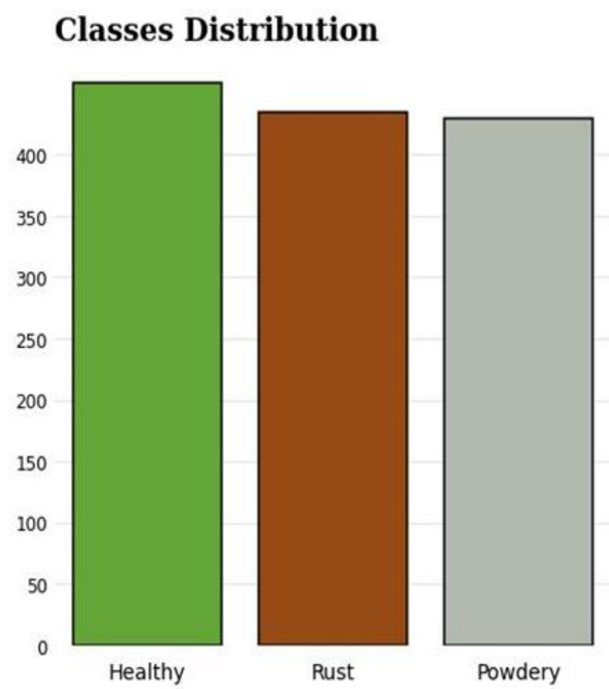


Figure 4.1: Dataset in class Distribution

In this experiment, we used a powerful CNN architecture through a transfer learning approach for the classification of plant diseases based on leaves. The training parameters listed in (Figure 4.1), were used to train the CNN models discussed in Section 3.7. These parameters provide the best outcomes during training after extensive testing. The effects supplied on this segment are associated with training with the entire database containing raw photos. As it's recognized as convolutional networks and that are capable of research capabilities whilst skilled on large datasets, effects completed whilst skilled with most effective photographs will now no longer be explored. By the convolution neural network it was taught to recognize real object detection. We used a prevailing CNN architecture, Resnet35 to examine the process clearly and tried to find the best way to find accuracy. Our model came out with the best accuracy of 94% using raw images. The graph of the accuracy of the train and validation of our dataset is shown below (figure). In (figure), the graph of the train and validation losses is shown.

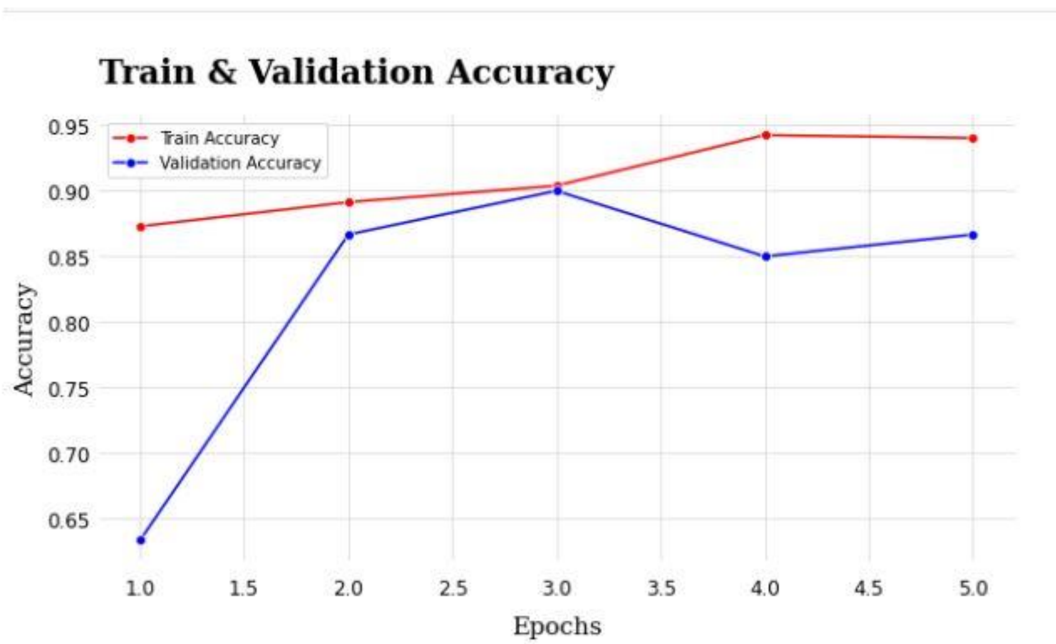


Figure 4.2 Model Train and Validation Accuracy Curve

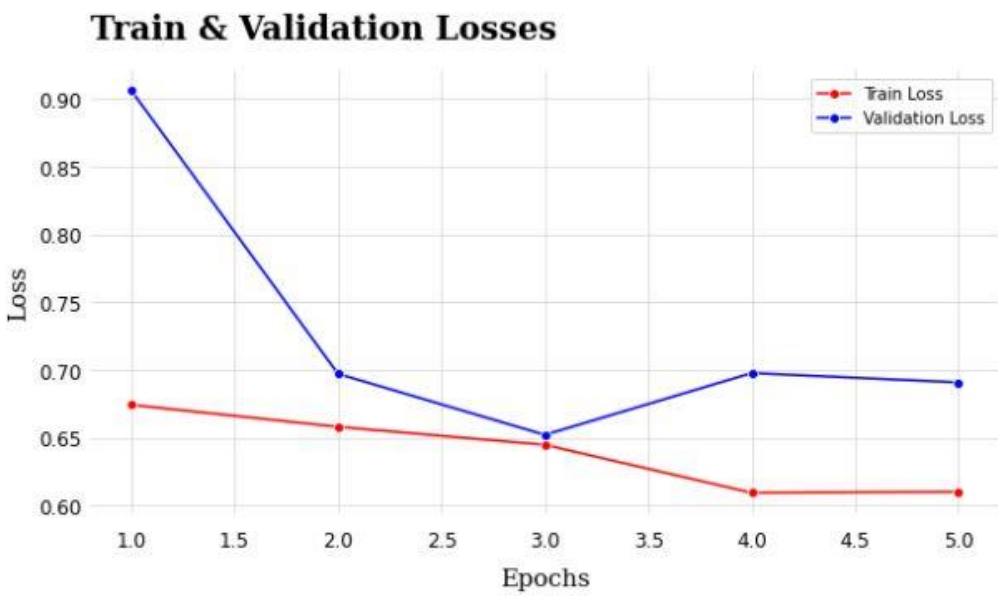


Figure-4.3 Model train and validation Loss Curve

On the accuracy curve it shows the accuracy of the train and validation dataset. On the other hand, in the train and validation loss curve it shows the losses quantity. The accuracy of feature extraction parameters improves as the number of iterations rises.

After setting up the model, several hyper parameters are set. These hyper parameters are set for training and to give a result accordingly. The data is kept in the first layer of the network while training the neural network. It also checks whether each neuron is equal to the label or not. After that the weight of the neuron is updated by back-propagation according to that and this process will continue till the new ability can be learned from the data in our training set. We need to have an idea of the confusion matrix so that we can establish these indexes particularly well. The confusion matrix will show the result produced by our model in binary classification and it will tell whether it is correct or not. There are 4 attributes in the confusion matrix and these are TP(true positive), TN(true negative), FP(false positive) and FN(false negative). Accuracy, precision, recall, and f1 is calculated as follows,

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

Precision predicts the proportion of correct predictions among all the positives predicted by the model. The precision can be calculated as follow,

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

Recall predicts the correct proportion of positives among all real positives[23]. The recall can be calculated by the following equation,

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1} = 2 \times \frac{precision \times recall}{precision + recall} \quad (6)$$

Precision correctly predicts the part of the model that has all the positive predictions. And recall correctly predicts the part of the model that has all the real positive predictions. F1 values are evaluated based on the confusion matrix.

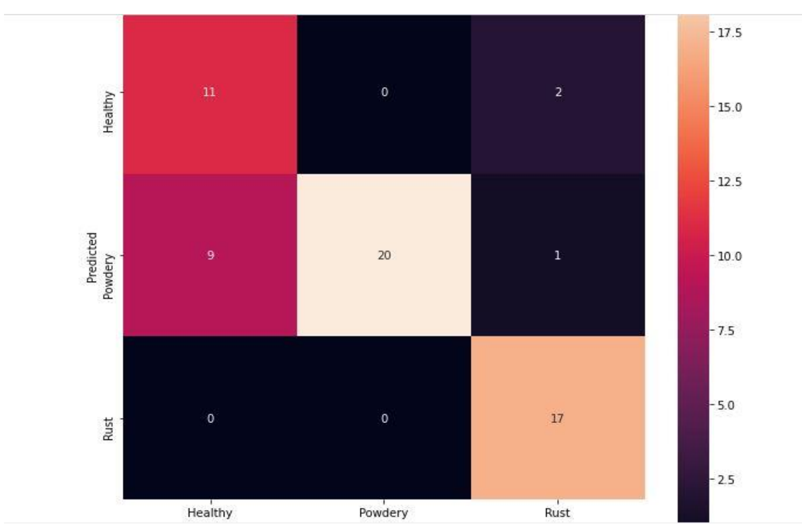


Figure 4.4: Confusion matrix of our model

The main target of our plant disease recognition work is to find the accuracy and train the model in such a way that we can get the highest accuracy. In that case, the more we can keep the value of accuracy, precision and recall, the more efficient our model will be to recognize plant leaf disease. On the other hand, the lower the value of f1, the more efficient our model will be to recognize plant leaf disease. Our trained model has a new capability after the training and analysis are complete and the capability is applied to new data.

Table 4.1 Classification Report

	precision	recall	f1 score	support
healthy	**0.93**	**0.76**	**0.84**	**50**
powdery	**0.77**	**0.96**	**0.86**	**50**
rust	**0.98**	**0.92**	**0.95**	**50**
Accuracy			**0.88**	**150**
Macro avg	**0.89**	**0.88**	**0.88**	**150**
Weighted avg	**0.89**	**0.88**	**0.88**	**150**

Table: 4.2 Diseases prediction by graph

S. NO	Leaf images	Accuracy	Loss	Disease Predicted
1		**0.9334**	**0.064**	**Healthy**
2		**0.896**	**0.896**	**Powdery**
3		**0.966**	**0.0966**	**Rust**

4.2 Comparative Analysis

There are several deep learning methods to identify the types of plant disease like ANN, CNN (Resnet34), etc. Compared to these, Resnet35 came out with the highest accuracy of 94% to detect if the sample picture of the leaf is healthy or not. And if it's not healthy, then the leaf is rusty or powdery. So our CNN model was utilized and achieved 94% of accuracy which is very satisfactory. This work has given us the highest accuracy compared with other papers. Some researchers have worked nearly close to our work. A comparison between our work and some previous work on several plant disease detection is exhibited on the Table below:

Table: 4.3 Comparisons with Some Previous Work

Reference	Work	Method	Accuracy
Subham Malvankar et al. [9]	Plant Disease Detection using Image Processing	Naive Bayes Classifier and KNN	65% accuracy in Naive Bayes Classifier 80% Accuracy in KNN
Ramcharan et al. [13]	A mobile-based deep learning model for cassava disease diagnosis	Single-Shot multibox(SSD) model with MobileNet detector and classifier	80.6% accuracy on images 70.4% accuracy on video
Mishra et al. [24]	Recognize two corn leaf disease (rust, northern leaf blight)	DCNN (Deep Convolutional Neural Network)	88.46%
Fuentes et al. [25]	Detect diseases and pests in tomato plants using images captured in-place by camera devices	VGGNet and Residual Network (ResNet)	83% (mean)

From the above table, we can see many researchers approach different types of algorithms and get various kinds of accuracy but out of all of the accuracy our model accuracy is the highest accuracy which is 94%. We used “Image Enhancement Technique” and CNN(Resnet34) and in our data preprocessing and also resize the images according to our requirement. This model can correctly identify rust, powdery and healthy leaves and it could be a far better solution for our farmers and their farming.

CHAPTER 5

Conclusion And Future Work

5.1 Conclusion

Plant diseases hamper production and quality of fruits, vegetables, fiber and biofuel crops. Farmers spend a lot of money on plant disease control, often without sufficient technical support and that results in poor disease control and harmful results. However, the knowledge of identity and multiplicity of plant leaf diseases occurring in Bangladesh are largely lacking. So, it is a foremost requirement to detect plant leaf diseases from the beginning state of the diseases. We used a deep learning technique called CNN in our experiment, plant leaf disease recognition. We have used the CNN methodology for our work here and summarized the required DL principles. Then, applied three layers in our model. We used the Resnet35 model and got 94% of accuracy which is satisfactory enough. As we have successfully and correctly classified healthy, rust and powdery leaves in this experiment, in the future we will have built an intelligent system with which farmers can easily identify plant diseases and get direction accordingly to cure the disease at an early stage.

5.2 Future Work

We will be making a more exact and vigorous model and will develop for upgrading the precision of the plant-disease-detection framework based on leaves. A smartphone application will be created similarly later to identify plant diseases and show the data to the farmers. The application will make a significant contribution to the farmers to detect the plant disease as soon as possible and reduce plant diseases.

REFERENCES

[1] Hassan,S.M.; Maji, A.K.; Jasi ´nski, M.; Leonowicz, Z.; Jasi ´nska, "Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach", 9 June 2021.

[2] Chen, J.; Chen, J.; Zhang, D.; Sun, Y.; Nanehkaran, Y.A. Using deep transfer learning for image-based plant disease identification.

[3] Google Slides - Plant Life CyclesRust - Google Arts & Culture

[4] Atabay, H. A. 2016b. A convolutional neural network with a new architecture applied on leaf classification. IIOAB J 7(5):226–331

[5] Mokhtar, U.; Ali, M.A.; Hassanien, A.E.; Hefny, H. ‘'Identifying two of tomatoes leaf viruses using support vector machine. In Information Systems Design and Intelligent Applications’’

[6] Jeon, Wang-Su, and Sang-Yong Rhee; ‘’Plant leaf recognition using a convolution neural network.’’

[7] Liu, Bin, "Identification of apple leaf diseases based on deep convolutional neural networks.

[8] K.Padmavathi, and K.Thangadurai. “Implementation of RGB and Gray scale images in plant leaves

disease detection –comparative study”

[9] Shubham Malvankar, Bilwa Prabhu Moye, Riya Shirodkar, Vipul Soman, Nora Naik Saurabh Naik “Plant Disease Detection using Image Processing”

[10] Ferentinos, K.P. ‘’Deep learning models for plant disease detection and diagnosis.’’

[11] Sladojevic, S.; Arsenovic, M.; Anderla, A.; Culibrk, D.; Stefanovic, D. Deep neural networks based recognition of plant diseases by leaf image classification. Comput. Intell. Neurosci. 2016, 2016, 1–10.

[12] Ramcharan, A.; Baranowski, K.; McCloskey, P.; Ahmed, B.; Legg, J.; Hughes, D.P. Deep Learning for Image-Based Cassava Disease Detection. Front. Plant Sci. 2017, 8, 1852.

[13] Ramcharan, A.; McCloskey, P.; Baranowski, K.; Mbilinyi, N.; Mrisho, L.; Ndalahwa, M.; Legg, J.; Hughes, D.P. A mobile-based deep learning model for cassava disease diagnosis. Front. Plant Sci. 2019, 10, 272.

[14] Geetharamani, G.; Pandian, A. Identification of plant leaf diseases using a nine-layer deep convolutional neural network.

[15] Md. Tarek Habib, Md. Jueal Mia, Mohammad Shorif Uddin, Farruk Ahmed "An in-depth exploration of automated jackfruit disease recognition". In Journal of King Saud University –Computer and Information Sciences, 25 April 2020.

[16] Dataset, available at [https://drive.google.com/drive/u/1/folders/1mgRrS7VSwFXhqgzYk2xK3T7wMh3Ay3K7]

[17] kaggle, available at [https://www.kaggle.com/rashikrahmanpritom/plant-disease-recognition-dataset]

[18] Yu, J.; Sharpe, S.M.; Schumann, A.W.; Boyd, N.S. 'Deep learning for image-based weed detection in turfgrass'. Eur. J. Agron. 2019, 104, 78–84.

[19] Guan, S.; Kamona, N.; "Loew, M. Segmentation of Thermal Breast Images Using Convolutional and Deconvolutional Neural Networks." In Proceedings of the 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 9–11 October 2018

[20] Fakhry, A.; Zeng, T.; Ji, S. "Residual Deconvolutional Networks for Brain Electron Microscopy Image Segmentation." IEEE Trans. Med. Imaging 2017, 36, 447–456.

[21] Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Wang, G. "Recent Advances in Convolutional Neural Networks. Pattern Recognit." 2015, 77, 354–377.

[22] Machine Learning Mastery, available at [A Gentle Introduction to Transfer Learning for Deep Learning], last accessed on 30-10-21 at 11.00 PM

[23] Fuentes, A.; Yoon, S.; Kim, S.C.; Park, D.S. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. Sensors 2017, 17, 2022.

PLAGIARISM REPORT

Plagiarism Checked by **Md. Zahid Hasan,** Assistant Professor**,** Department of CSE Daffodil International University	08-12-21

Plant Leaf Disease Recognition Using Convolutional Neural Network

ORIGINALITY REPORT

29%	23%	16%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
2	Submitted to Manchester Metropolitan University Student Paper	3%
3	www.mdpi.com Internet Source	3%
4	Submitted to Daffodil International University Student Paper	2%
5	irjmets.com Internet Source	2%
6	turcomat.org Internet Source	1%
7	"Smart Systems: Innovations in Computing", Springer Science and Business Media LLC, 2022 Publication	1%