Tea leaf Disease Identification Using Machine learning Approach

BY

MARUF HASAN MREDUL
ID: 191-15-12686

AND

Md. Shakil Khan
ID: 191-15-12205

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

Supervised By

Mr. Dewan Mamun Raza
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Dr. Sheak Rashed Haider Noori
Professor & Associate Head
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

24 JANUARY 2023

APPROVAL

This Project/internship titled **"Tea leaf Disease Identification Using Machine learning Approach"** submitted by MARUF HASAN MREDUL, Bearing Student ID: **191-15-12686**, Md. Shakil Khan, Bearing Student ID: **191-15-12205,** to the Department of ComputerScience 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 **24 January2023.**

BOARD OF EXAMINERS

24/01/23

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Shafin 24.01.23

Dr. Mohammad Shamsul Arefin
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Md. Sabab Zulfiker
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

24.1.2023

External Examine

Dr. Ahmed Wasif Reza
Associate Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Dewan Mamun Raza**, Lecturer, 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:

Mr. Dewan Mamun Raza
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Dr. Sheak Rashed Haider Noori
Professor
Department of CSE
Daffodil International University

Submitted by:

MARUF HASAN MREDUL
ID: 191-15-12686
Department of CSE
Daffodil International University

Md. Shakil Khan
ID: 191-15-12205
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First of all, I would like to express my heartfelt thanks and thanks to Almighty God for the divine blessing that allowed me to successfully complete last year's project.

We are really grateful and wish our profound indebtedness to **Mr. Dewan Mamun Raza, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge &keen interest of our supervisor in the field of ―*Machine Learning*‖ to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

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

We would also like to express our sincere thanks to the other members and staff of Ispahani tea Estate Limited and Finlay Tea Estate and BTRI (Bangladesh Tea Research Institute) for their kind help in completing our project.

At the end of the course we would like to thank all my teammate from Daffodil International University who participated in this discussion. Lastly, we must respect the unwavering support of our parents and the hearts of our patients.

ABSTRACT

Agriculture remains the most important sector of Bangladesh economy, contributing 19.6 percent to the national GDP and providing employment for 63 percent of the population. Agriculture is the largest livelihood provider in Bangladesh. Most of all, for the vast rural population. Tea is one of the important crops in Bangladesh. It contributes 1 percent of total countries GDP. And it contributes 3 percent of global tea production. The tea industry employs one and a half lakh workers. Bangladesh tea industry now crossing BDT 4000 crore. But every year the tea industry faces a large amount of losses due to leaf diseases like bacterial, gray blight or fungal infections. For those diseases and fungal infection tea production decreases. Identifying diseases from tea leaf is critical and challenging work. But we have to do that to sustain tea demand worldwide. And it is also important for Bangladesh economy. Now tea worker finds those diseases by his eyes. Most of the time affected tea remains in the garden and it spread diseases tree to tree. It decreases tea production and tea quality. We need a solution for that. So, our purpose of this project is to develop a system that can detect those tea leaf diseases. We will do that by using machine learning techniques like image processing. In image processing there are some models like SVM, and CNN (convolutional neural network) based models are usually used for research. We chose SVM, CNN and VGG-16(CNN-based model) to identify tea leaf disease. Because we previously see those pre trained models work find in this kind of image. We train those models of our 75 percent of data, and we keep 25 percent data for our testing purpose. After finishing training when we tested our models, we got 98 percent accuracy rates from the model VGG-16.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES **PAGE**

LIST OF TABLES

TABLES **PAGE NO**

CHAPTER 1
INTRODUCTION

1.1Introduction

Tea was invented in 2737 BC accidentally. Shen Nung the Chinese emperor and his servant boiling water suddenly some leaves drop into the boiling water. They drink the boiled water and feel energetic. Because tea have amino acid and it produce a calming effect. Science then tea is a common beverage drink by people worldwide and it easy to make. In boiling water, we have to put some processed tea leaf and wait for around three to five minutes, our tea is ready. Tea is also used as a medicinal beverage. Tea has vitamins E, C, A and B. In Bangladesh specially on winter season when we get fever, we drink lots of tea for cure. Also, tea is the first thing that most Bangladeshis drink in the morning. Tea produced mostly in Asian countries. Bangladesh is one of them. Bangladesh produces a significant amount of tea. Bangladesh produced tree more than 180 years. In Bangladesh there are around 2, 81,107 acres of land where tea produced. In Bangladesh, we have around 172 tea garden/state. And the tea-producing places are Srimangal, Sylhet, Habiganj, Chittagong, MaulviBazar, Brahmanbaria, Panchagarh, Rangamati. Among them Srimangal is the tea capital of Bangladesh. Sylhet is also known for good quality tea producer. Bangladesh tea export is the second largest among all other exports crops. And if we look after the amount, it is 1 percent of our country's GDP. But every tea season tea producers face huge number of losses for tea leaf diseases and all the 100 percent 20 to 30 percent tea leaf and tea are affected from those diseases. And some of diseases are spread tea leaf to tea leaf. Farmer so far separated affected tea leaf by measuring his own eye, sometime the affected tea left among the healthy tea, and it spread diseases. It is hard to detect leaf diseases by eyes. And it takes time and money. So, we can solve this problem by applying the Machine Learning Algorithm.

1.2 Motivation

Tea is essential for our daily life, and it is beneficial for our healthy life. Tea has vitamins E, C, A and B. And those are the most important vitamins for the human body that are found in tea. Tea leaf disease is an exquisite problem for our tea export. In Bangladesh most of the farmers are in rural areas, they do not have the proper guideline how to deal with tea diseases and therefore the farmers are careless of tea leaf diseases, and it is major concern because it is not only reducing the tea quantity but also it reduces tea quality. And therefore, tea producers face a huge amount of loss in every tea season. It is necessary to be concerned about leaf disease to grow and sustain our tea industry. For that reason, we will do this work.

1.2Objective

Our main objectives are,

- Build a machine learning model that will work fine in Bangladeshi tea leaf diseases.
- Build a machine learning model that will reduce time, money and effort in detecting tea leaf diseases.
- Our machine learning model will be so easy that anyone, especially the tea farmer, will easily use that.

1.4 Research Question

Tea leaf disease detection is very difficult work to do. In start of the work, we don't know what to do, how do we start our work. In that time, we have some question. Those are,

- Which machine learning algorithm is fine for our work.
- How much data do we need for our project.
- Where to collect those data

1.5 Expected Outcome

As a result of this work, we will develop a system that will help the tea producer of Bangladesh as well as other tea producing country. And it will help tea producers reduce time, money and overall working efficiency. We have some objectives in the work's conclusion. Some of those are,

- Reduce the amount of time it takes to find tea leaf illness.
- Reduce the price of disease detection.
- Increase the amount of tea produced.
- Accurately classify each image of a leaf disease.

As a result of this work, we will develop a system that will help the tea producer of Bangladesh as well as other tea producing country. And it will help tea producers reduce time, money and overall working efficiency. We have some objectives in the work's conclusion. Some of those are,

- Reduce the amount of time it takes to find tea leaf illness.
- Reduce the price of disease detection.
- Increase the amount of tea produced.
- Accurately classify each image of a leaf disease.

1.6 Layout of the Report

In chapter 1, we have discussed why we are going to do this project. What are the necessary doing this project. Why we are so motivated about this project. Some question what we face earlier and the report layout. In chapter 2, we have find out what other people do the same kind of work or not. Summary of Our research paper. Some challenge what we are facing to do this work. In chapter 3, we have discussed the data collection procedure, data preprocessing and what are The models that we use for our work and why we should choose this not the other models. In chapter 4, we have discussed how to setup for the experiment and the result of it. And The discussion.

In chapter 5, we have discussed about how our research will impact in the society and Environment. The Ethical aspects of our research. In chapter 6, we have a summary of the research, some suggestions for future research, and our findings were presented in the fifth chapter. We also talk about some of our limitations and problems, but this chapter ends by showing how much the area has potential for the future. The distribution of the comprehensive project summary of our study research is the responsibility of this chapter.

CHAPTER 2
BACKGROUND

2.1 Introduction

Problems in computer vision and image processing are being solved by machine learning. Machine learning can solve that kind of problem better than any of the others. But it is hard to understand machine learning insider algorithm. The purpose of this project is to develop machine learning models who can work well with tea leaf diseases.

2.2 Related Works

Convolutional Neural Network (CNN) based models work very well in image processing. And the common use of it is tree leaf diseases detection. Some work that has been done with this tree leaf diseases detection are, one of the works is tea leaf diseases detection. In this they identify three types of diseases they are, brown blight, blister blight and algal leaf spot. They use faster R-CNN for classifying those images. Their accuracy was about 69.79%. And they apply it on 2437 leaf images dataset [1]. Another paper is tomato leaf diseases detection. In this paper they classify 10 different types of leaf images, and they model they used for SVM and HOG. They used 3000 tomato leaf images, and their accuracy was about 97% [2]. Another paper is about leaf disease detection using deep CNN(DCNN). They classify the leaf image into 37 separate classes [3]. Another paper is potato leaf disease classification. In this paper they apply k-means clustering algorithm to identify unhealthy leaves [4]. Another paper is about plant disease detection.in this paper 87848 images to train and test their model based on CNN model and they got 96.53% accuracy [5]. Another paper they used VGG-16 transfer learning algorithm. They fine tune the VGG-16 architecture, and they got 98.69% accuracy [6]. Another paper is about skin disease detection in this paper they used Logistic regression Classifier [7]. Another paper is for brain tumor detection. In this paper they used VGG-16 algorithm and they got 96% accuracy [8]. Another paper is about detecting leaf disease. In this paper they used CNN architecture. They apply it around 2000 leaf images and they got 92.78% accuracy [9]. Another paper they used VGG-16 model to identify tea leaf disease. They got 98.68% accuracy [10]. In our model we as well as used the VGG-16 model to identify tea leaf disease. We see most of the work done here about leaf disease detection. Some of them also in tea leaf disease detection about the number of paper and considering our countries weather the work is only two and the work in not UpToDate. So,

2.3 Research Summary

The purpose of this project is to develop a machine learning image classification Convolutional Neural Network (CNN) model, for the classification of the tea leaf picture. Some years ago, ML (machine learning) techniques were first utilized for image-based categorization. There is lots of work done in this field. We can gain knowledge by looking at the background of ML (machine learning) related research. After testing, questioning, analyzing, and comparing all relevant studies on tea leaf disease detection, we arrived at a variety of results. But most of the work done is based on other countries' tea leaf. A very few works done in based on Bangladesh tea leaf but those are too old, and the disease may also different now and the important thing the ML (Machine learning) techniques they used that is also old. Another factor is accuracy rate; most projects accomplish up to 92 to 96% of the accuracy rate. Now, we propose to do Tea leaf disease detection using updated ML (Machine Learning) technique with better accuracy.

2.4 Scope of the Problem

In every phase of our work, there are numerous issues. It is tough to determine difficulties before we begin our work. But we can do it as our work needs. However, the most Difficult aspect of our work is information assortment. We must collect images from the tea state located far away from the capital. And we have collected those in such a way that there is no background noise in the image. That is tough work to do.

2.5 Challenges

To finish our work, we needed to conquer a few hardships. The most difficult part of our work was collecting data. For collecting data, we first chose a tea leaf photographer. Because we need to our dataset as much as possible noise free. Then we should go to the tea state located on srimalgal, Syllet. And we collect the data Different tea state, like Ispahani tea state, finley tea state, BRTI tea state. And similar types of tea leaf are also a big problem for us. The next problem is to select a model that is suitable for our work. After that, in order to train the large number of images that we ultimately selected, a powerful GPU is required. We need a GPU because it takes much longer to train a model without a GPU than it does to train a model with

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

We will describe our entire working method in detail in chapter 3. Our method of operation has been broken down into several phases. To begin with, we'll go through information assortment and pre-handling. The most recent and widely used method for image classification, the Conventional Neural Network (CNN), was then the topic of our discussion. The input and output process, as well as the training and test sets, will be discussed in the concluding section. To make the idea clear, we'll give a perfect example.

3.2 Research Subject & Instrumentation

The subject of the study gives a clear picture of the area of our research. This study aims to distinguish between healthy and diseased tea leaf varieties. This section discusses our findings, selects our model, gathers the best data, prepares and trains our model, and finally puts our model to use. A Machine (Conventional Neural Network) learning model for image classification is presented by us. NumPy, matplotlib, and other Python data science libraries are also utilized by us. We work on computers running the Windows operating system. To carry out this experiment on any operating system, we will require some instruments. The instruments and requirements listed below are:

i) Machine Learning Models
ii) Python Language
iii) Google Collaboratory.

We selected Python because it is free and easy to use, and it can be used for sophisticated algorithm readability in machine learning applications. In addition, we use Google Collaboratory, a free online cloud based Jupiter notebook environment that we can finish our project.

3.3 Data Collection Procedure

The detection of tea leaf disease based on the Bangladeshi tea plant is our objective for this study. Therefore, we require a recently collected dataset from the tea state or garden. We need to create a brand-new, authentic collection of images of tea leaf disease for our research. Therefore, we decided to visit some of the tea states in Sylhet, Sremangal. For the purpose of collecting images of tea leaves, we traveled to the states of Ispahani, Finley, and BRTI.

We settled on a selection of tea leaf images; Some of them are diseased, while others are healthy. Each image has a different resolution and is saved in JPG format. There are approximately two thousand images in total. The leaf on the right bellow talbe is in good health, while the leaf on the left is ill.

There are around 2,000 tea leaf pictures in our assortment. Approximately 1000 of the 2000 images depict healthy tea leaves, while approximately 1100 depict diseased leaves. We used 25% of the 2000 data for testing, while 75% of the data was used for training. The gathering of those tea leaf data was the first step in the development of our thesis. We simply gathered the tea states' raw data.

In the bellow dataset there are two types of leaf one is healthy tea leaf another one is diseases tea leaf. The diseases dataset has four category bacterial, gray and blight and the fungal infection.

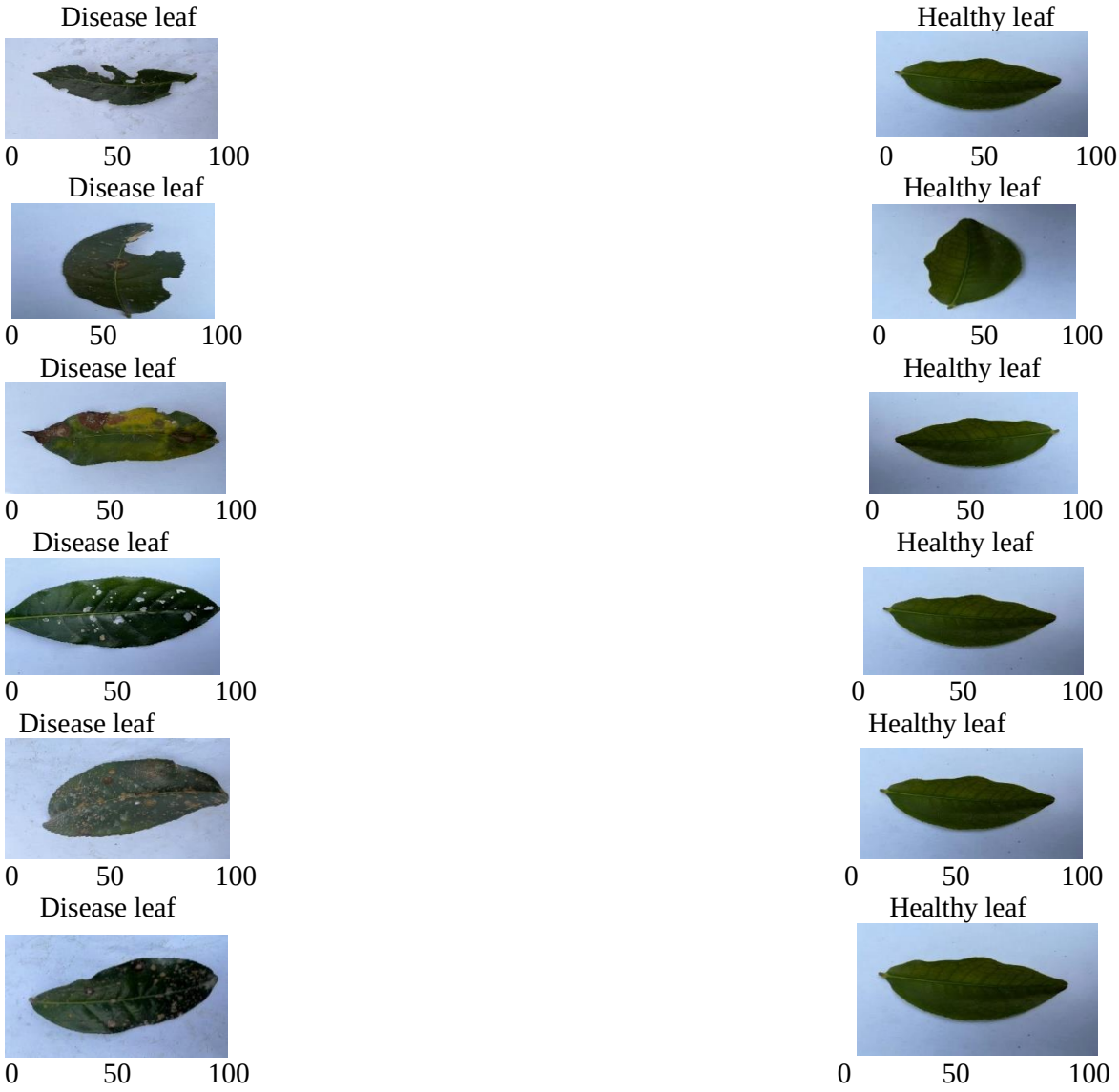


Figure 1.0: A tiny portion of the initial dataset

3.4 Proposed Methodology

The entire work is broken up into a number of sub-processes to ensure that the model is correctly identified. Figure shows the complete procedure.

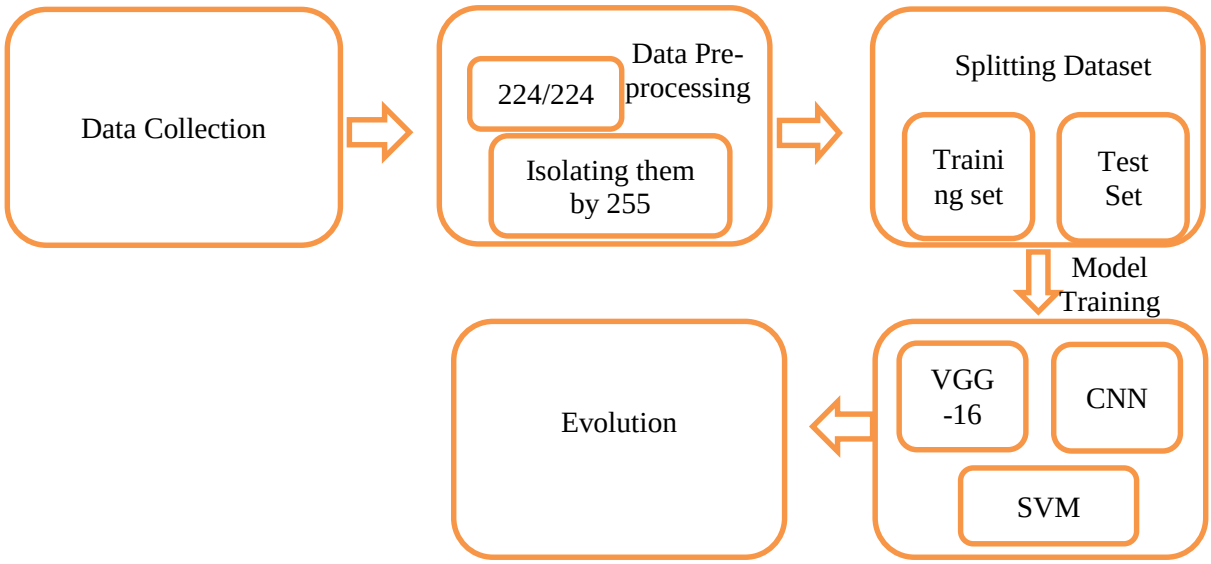


Figure (1): Schematic of the proposed method

As we have such countless various shapes and sizes of leaf picture, we should initially accumulate the information and afterward preprocess the dataset. After that, we split our dataset into two sets—a test set and a train set. Then, some models from machine learning are used. In this instance, we employ three distinct varieties of machine learning algorithms. These are the main ones:

i) CNN

ii) SVM

iii) VGG-16

For all our models we use the same parameters. We first categories our dataset into main to division one is healthy tea leaf another one is affected by disease tea leaf. Then the affected tea leaf disease we categorize into three main subcategories of the diseases. Then we Split them into 75 and 25 percent. 75 percent for taring our machine learning models and the rest 25 percent is for testing our models.

3.5 Data Preprocessing

We begin by gathering information from the srimangals some of tea state. However, our data became disorganized as we clicked the tea leaf image at random. Tea leaf images come in a variety of sizes and resolutions. Therefore, prior to submitting our datasets to the training of machine learning models, we should prepare them. Data processing is essential for model training and accuracy enhancement because, among other things, it reduces computer costs. Our pictures have been resized to 150 by 150 pixels. In order to avoid overfitting our model, the data are then shuffled. Furthermore, we divided our leaf images into three distinct files. We diminish RGB values by isolating them by 255 to standardize our information.

3.6 Convolutional Neural Network

A convolutional neural network (CNN) is a feed-forward neural network that typically processes data in a grid-like topology for the purpose of analyzing visual images. It is also referred to as a Conv-Net. An image's objects are identified and categorized using a convolutional neural network. A deep learning algorithm called Convolutional Neural Network (CNN) can use images as input and apply learnable weights and biases to different objects in the image to distinguish them from one another. In most cases, inputs for pattern recognition are necessary for the operation of a CNN. A common type of machine learning network known as CNN is utilized to address issues with image visualization such as classification. Finally, the final output is obtained when all linked layers establish all connections to the preceding layer.

The convolution layers equations are $x\ell ij=\sum a=0m-1\sum b=0m-1\omega aby\ell-1(i+a)(j+b)$.
Here N is the number of layers m and ω is filter.

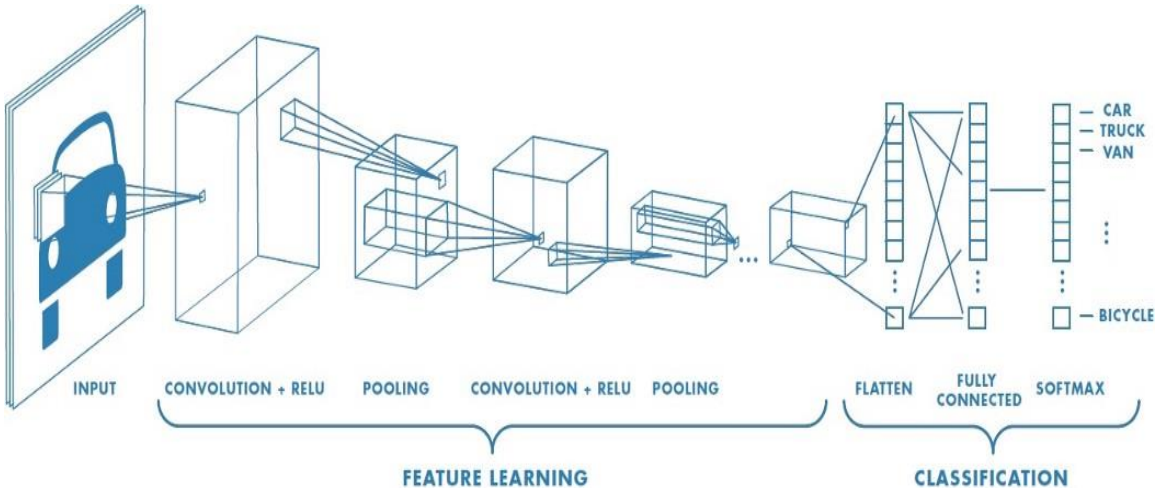


Figure (2): A typical architecture for a CNN

We use a ML procedure in this review to classify photographs of unhealthy and solid leaves. In our study we used pre-defied machine learning algorithm because it tough and not efficient to develop a machine learning model from scratch. And pre-define machine learning models work very well. VGG-16, SVM, and CNN are the three machine learning models we use in this study.

VGG16: A convolutional neural network is a feed-forward neural network that typically processes data in a grid-like topology and is used to analyze visual images. It is also referred to as a Conv-Net. An image's objects are identified and categorized using a convolutional neural network.

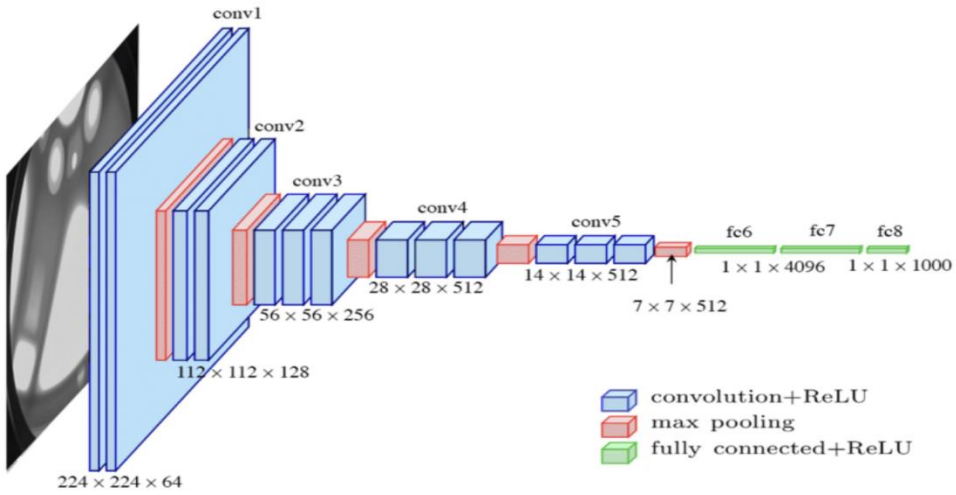


Figure (3): Architecture of VGG16

In the diagram that came before it, the non-linear activation function, which is a rectified linear unit, and all of the blue rectangles serve as representations of the convolution layers. Starting with a channel size of 64, this design gradually increases by a factor of 2 after each max-pooling layer until it reaches 512. The architecture is easy to understand. By changing the final SoftMax dense layer unit to any number we choose based on the classes we need to categorize, we can make this model work for any number of classes.

SVM: The classification algorithm SVM is very good. It is a supervised learning algorithm primarily utilized for data classification. Using a set of label data, SVM trains. The fact that SVM can be used for both classification and regression problems is its main advantage.

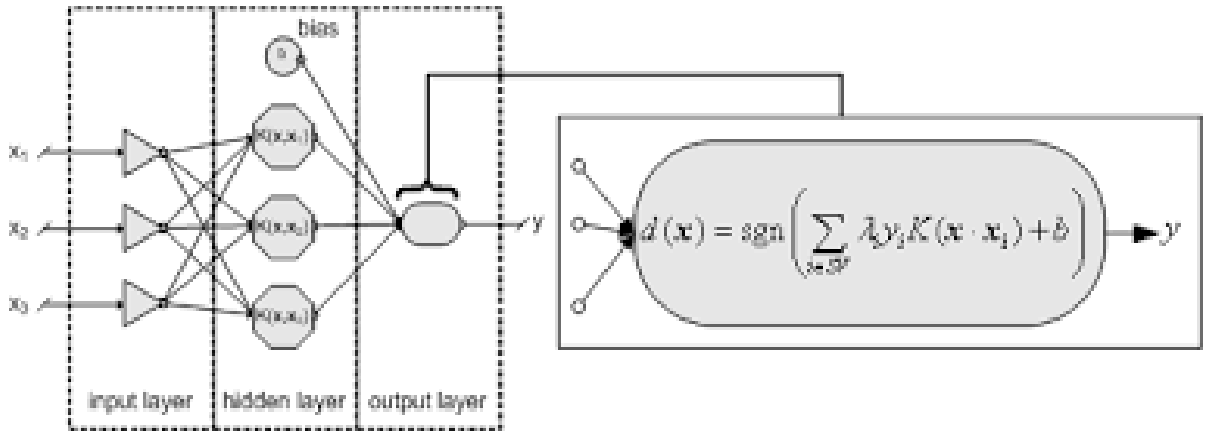


Figure (4): SVM architecture

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used to solve regression or classification problems. However, most classification problems make use of it. We plot each data point as a point in n-dimensional space using this SVM algorithm, where n is the number of features you have and the value of each feature is the value of a specific coordinate.

CNN: A Convolutional Neural Network (CNN) is a type of machine learning algorithm that can take in an image as input, assign weights and biases to various aspects or objects (learnable weights and biases), and distinguish between them. When compared to other classification algorithms, the amount of pre-processing required by a CNN is significantly lower. CNN is able to learn these filters and characteristics with sufficient training, whereas in primitive methods, filters are hand-engineered.

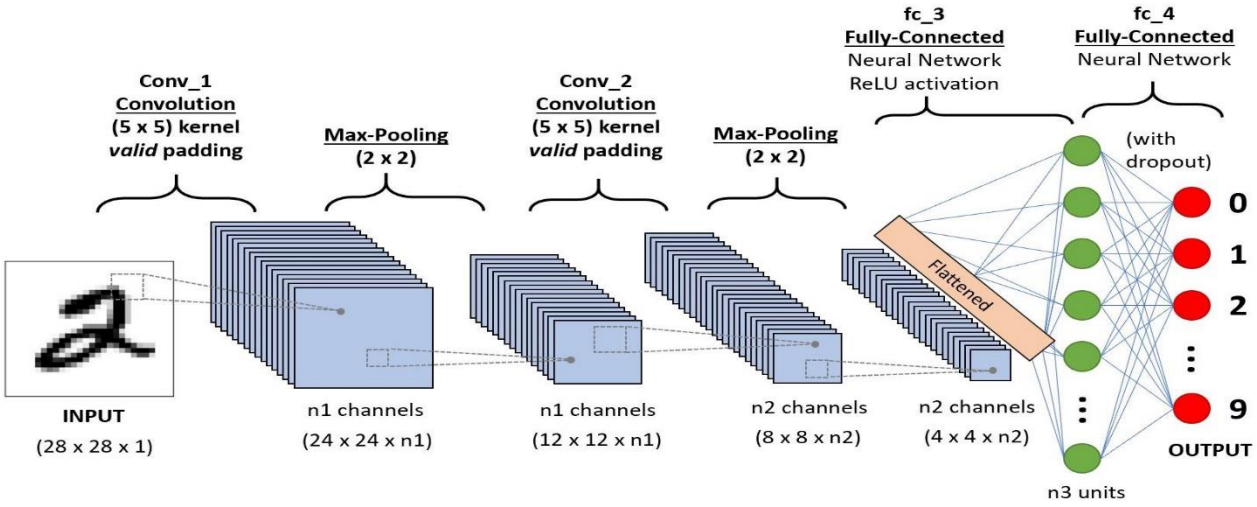


Figure (5): Architecture of CNN

The organization of the Visual Cortex served as inspiration for the architecture of CNN, which is comparable to the connectivity pattern of neurons in the human brain. Individual neurons

answer improvements just in a limited locale of the visual field known as the Responsive Field. The entire visual area is covered by a collection of such fields that overlap.

3.7 Train the model

We are now prepared to train our model, which uses 75% of our entire tea leaf dataset, after pre-processing it and creating the train and test datasets. We also used a label encoder to reduce the loss prior to fitting the data into the model. The model is then trained for up to 100 epochs.

3.8 Implementation Requirements

Top Complete our research we have to train and test our models. For that we need a environment. The bellow is our system requirement to train and test our model.

- Operating system (Windows 7 or later or any other up-to-date operating system)
- RAM (minimum 4 GB; recommended)
- Hard Disk (minimum 500 GB)
- GPU (recommended) In addition, the following development tools should be utilized:
- Collaboration with Google
- Python library
- Python Environment

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSION

4.1 Introduction

We investigate the overall performance evaluation of our model in Chapter 4. The number of parameters, level of precision, and other details will also be discussed. Then, we'll look at various transfer learning models like CNN, SVM, and VGG16. For straightforward supporting, we use a few diagrams, designs, a disarray framework, and order reports.

4.2 Experimental Setup

In order to find the best model for the detection of tea leaf diseases, various models were presented. After each model was used, its accuracy, loss, precision, recall, and f1- score were looked at. To show how well each model performed, the confusion matrix was shown for each model. The accuracy, precision, recall, and f1-score are all calculated using the following formulas:

Accuracy = (TP+TN) / (TP+TN+FP+FN)

here, TP = True positive, TN = True Negative.
Where FN stands for False Negative, TP stands for True Positive, and FP stands for False Positive.

4.3 Experimental Results & Analysis

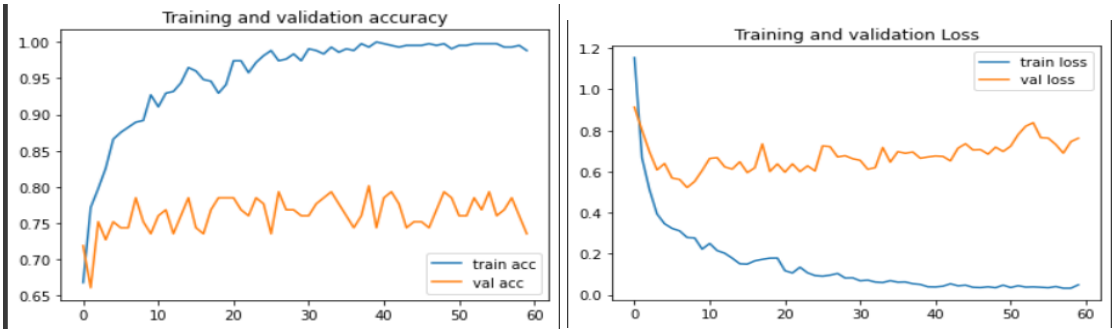
Bellow table are shown our three models loss and accuracy.

Models	Loss	Accuracy
VGG-16	09%	98%
CNN	41%	82%
SVM	33%	33%

Before applying those machine learning algorithms, we do preprocess our data. We have around 2000 tea leaf images. First, we split our dataset into 75 percent and 25 percent of the data. Then we go preprocess we resize our data in 150 by 150 format then we isolate it by 255. Then we fit those data into our models and we got the above accuracy.

VGG16

Figure (vi): (i) shows the model's accuracy during training and testing, whereas (ii) shows the model's loss during training and testing.



(i) (ii)

Figure: (6): (i) VGG16 model precision (ii) Loss of VGG16 model

Here we can see in VGG-16 model we have very higher accuracy compair to the loss. And if we See the loss, losss is bellow 1 percent throughout the whole fiting time.

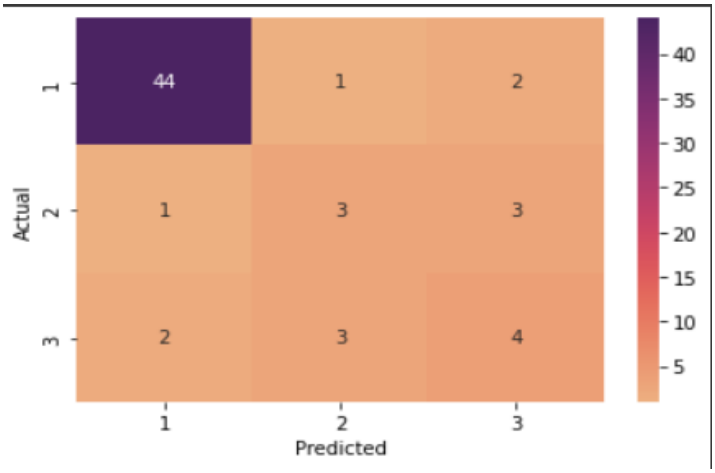
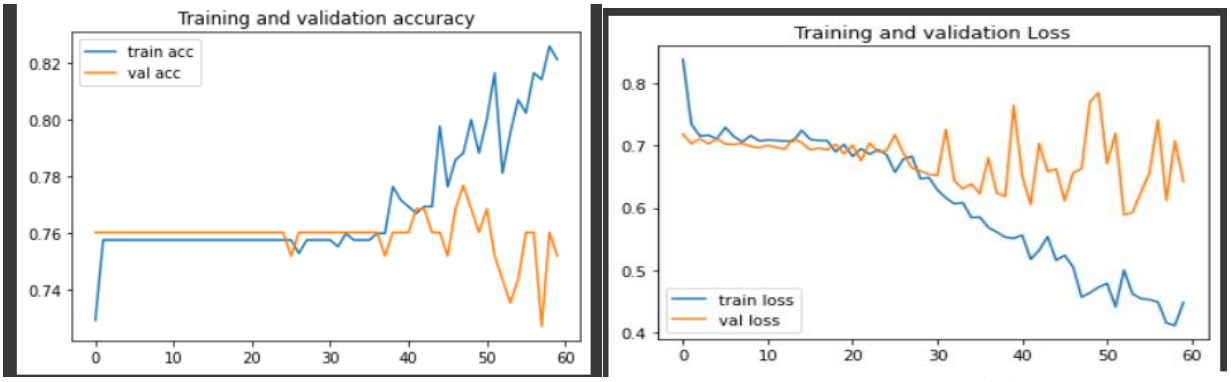


Figure 4.2 depicts the confusion matrix for the VGG16 model.

From the confusion matrix we can also see our data is very good so our true positive rate is high.

CNN

Figure (viii): shows the model's accuracy during training and testing in (i) and its loss during testing in (ii).



(i) (ii)

Figure: (8): (i) Accuracy of the CNN model (ii) Loss of CNN model

Here we have got a fluctuation in first accuracy was good but the process go through the value for loss is high.

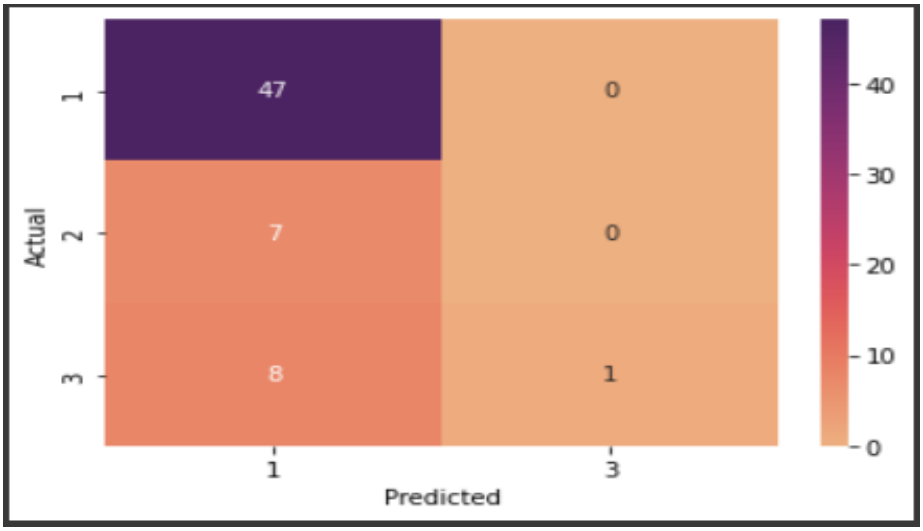
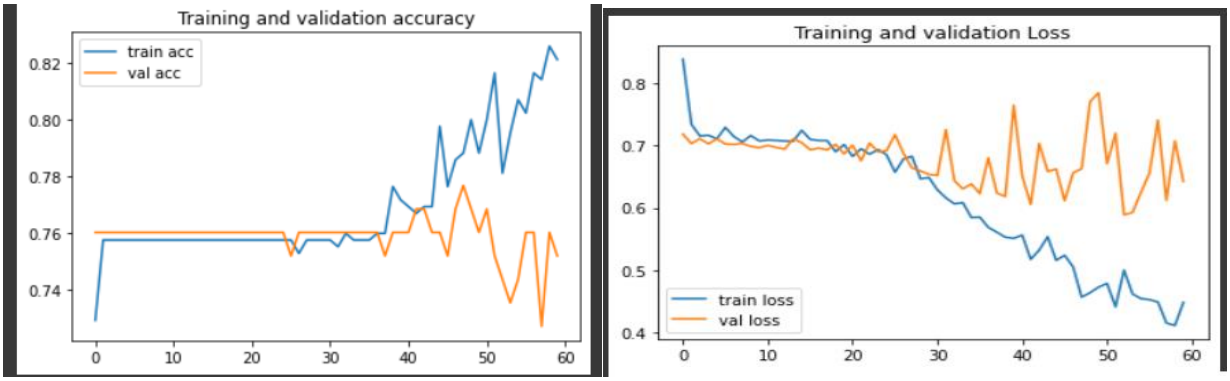


Figure: 9: Model confusion matrix for the CNN

Here we can also see the fluctuation between true positive and true negative.

SVM

Figure (ix) shows the model's accuracy during training and testing in (i) and its loss during testing in (ii).



(i) (ii)

Figure: (10): (i) VGG16 model precision (ii) Loss of VGG16 model

Here we have got a fluctuation in first accuracy was good but the process go through the value for loss is high.

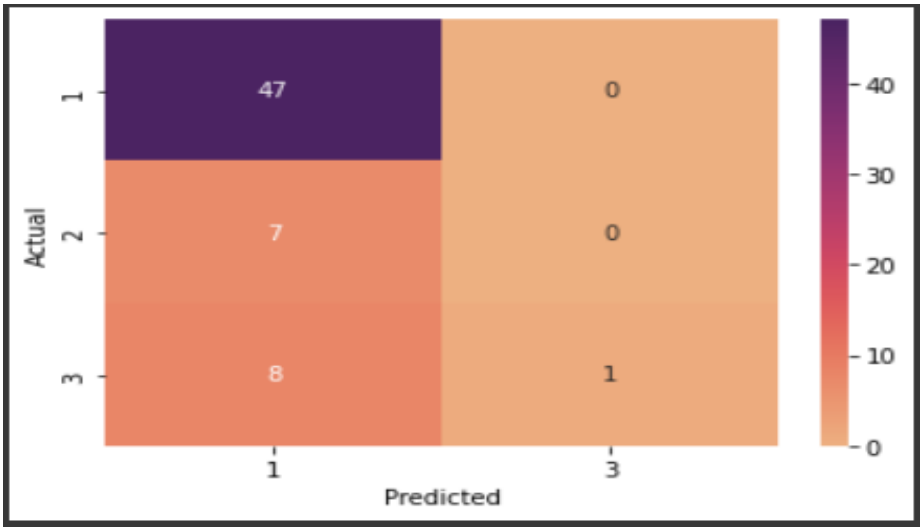


Figure: (11): Model confusion matrix for the SVM

Here we can also see the fluctuation between true positive and true negative.

4.4 Discussion

To compare the accuracy of various architectures, we created a chart.

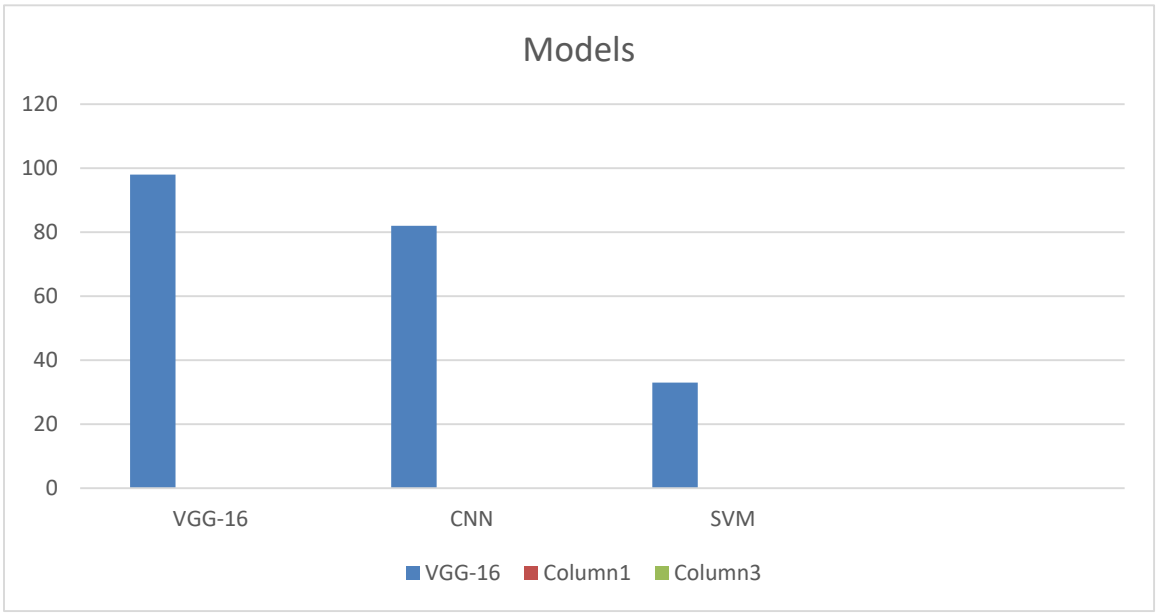


Figure: (12): Different models have different accuracy.

The accuracy of VGG-16 is 98%, that of CNN is 82%, and that of SVM is 33%, as shown

in the graph above. We can see that VGG-16 produces the most accurate results when compared to other models' accuracy. Thus, our recommended model for distinguishing infection tea leaf from photographs is VGG-

CHAPTER 5

IMPACT ON SOCURITY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Our nation has numerous tea states and gardens. Some of them are grown as industrialized tea, while others are grown as individual tea. They should be commonplace tea leaf sickness identification. This study is intended for individuals and businesses that cannot always identify tea leaf diseases. These diseases cannot be detected in the majority of tea states. Consequently, they are losing a lot of money from their business. Because there are few experts in this field in Bangladesh and no specialist is available in the remote region to assist them. In this way, our venture might be capable help them by perceiving the tea leaf issues and forestalling a huge misfortune.

5.2 Impact on Environment

Most farmers are not aware of tea leaf diseases. And for that every year they face a significant amount of loss. And it is also the causes of garden pollution and farmer cut those tree and through away next to the garden it also affected to the other tree. By implementing our project, garden pollution can be reduced by farmers. We will be able to inform them of the actual situation and enable them to take action that will significantly reduce the use of hazardous chemicals thanks to our initiative.

5.3 Ethical Aspects

We should be ethical in every aspect of our life. In our case we physically go to tea garden in srimangal, Sylhet. And capture tea leaf disease and healthy image by our own hand. And we talk to a scientist of Ispahani tea state which diseases are those and based on his we can separate those tea leaves.

5.4 Sustainability Plan

In our research we got a very good accuracy from Conventional Neural Network (CNN) based model VGG-16. In the near future our plan is to develop an android app and an IOS app for more convenient use of those machine learning models. And also, we have a plan for developing apps that can work based on Bangla speech. Because most of the garden worker are not educated and they cannot read English.

CHAPTER 6

SUMMARY, CONCLUSION AND FUTURE WORKS

6.1 Summary of the Study

Since Bangladesh exports tea, one of the most pressing issues facing the tea industry is the identification of tea leaf disease. This can be accomplished in a number of ways. However, the majority of them are incapable of identifying leaf diseases. In this experiment, we used approximately 2000 images of tea leaves to distinguish between healthy and diseased leaves. To select the most suitable machine learning model for this endeavor, we made use of a pre-built machine learning algorithm. We chose the VGG-16 model because, when compared to the other models, it has the highest accuracy.

6.2 Future Works

After finishing this work now, we know how to train and test and get better accuracy in the image processing. So, now we can do other image processing work like some other tree leaf disease detection. And we can implement our knowledge to the field like Computer vision. And after finishing this work we also know which model can work fine with those kinds of data set. In our case we got higher accuracy from VGG-16 model.

6.3 Conclusions

This project has provided an opportunity for the people of Asia, particularly Bangladesh and the Indian subcontinent, to have a forecast of tea leaf disease detection that is more accurate because we collect the dataset from the tea state/garden for this research. despite the fact that the suggested VGG-16 model resulted in a test loss of 0.0901 points and an accuracy of 98%. We also talked about various graphs, the Confusion Matrix, and the classification report. We achieved a higher accuracy thanks to our use of a fundamental machine learning algorithm. The creation and improvement of a CNN model for the detection of leaf disease is the ultimate goal of this study. We will keep conducting research to enhance the system's efficiency

REFERENCES

[1] B. Liu, Y. Zhang, D. He, and Y. Li, "Identification of apple leaf diseases based on Deep Convolutional Neural Networks," Symmetry, vol. 10, no. 1, p. 11, 2018.

[2] Bakonyi M, Johnson C R (1995) The euclidian distance matrix completion problem. SIAM Journal on Matrix Analysis and Applications 16(2):646–654

[3] Dae Gwan Kim J Q, Bulanon D M (2009) Classification of grapefruit peel diseases using color texture feature analysis. International Journal of Agricultural & Biological Engineering 2(3):41–50

[4] G. Geetharamani and A. Pandian, "Identification of plant leaf diseases using a nine-layer deep Convolutional Neural Network," Computers & Electrical Engineering, vol. 76, pp. 323–338, 2019.

[5] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[6] 2. Dheeb Al Bashish M B, Bani-Ahmad S (2011) Detection and classification of leaf diseases using k-means-based segmentation and neural-networks-based classification. Inf Technol J 10(2):267– 275

[7] 1. Sulistyo, S.B.; Wu, D.; Woo, W.L.; Dlay, S.S.; Gao, B. Computational Deep Intelligence Vision Sensing for Nutrient Content Estimation in Agricultural Automation. IEEE Trans. Autom. Sci. Eng. 2018, 15, 1243–1257. [Google Scholar] [CrossRef]

[8] A. T. Sapkal and U. V. Kulkarni, "Comparative study of leaf disease diagnosis system using texture features and deep learning features," International Journal of Applied Engineering Research, vol. 13, no. 19, pp. 14 334–14 340, 2018.

[9] Shaveta Malik, Tapas Kumar, Amiya Kumar Sahoo, ―Image processing techniques for identification of fish disease,‖ 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP), Vol. 15, No. 5, August 2017.

[10] Yang, N.; Yuan, M.F.; Wang, P.; Zhang, R.B.; Sun, J.; Mao, H.P. Tea Diseases Detection Based on Fast Infrared Thermal Image Processing Technology. J. Sci. Food Agric. 2019. [Google Scholar] [CrossRef] [PubMed]

[11] Arshia Rehman, Saeeda Naz, Muhammad Imran Razzak, Faiza Akram and Muhammad Imran, ―A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning‖, Circuits Systems and Signal Processing (Circ Syst Signal Process), DOI: 10.1007/s00034-019-01246-3.

[12] Jeong-Seon Park, Myung-Joo Oh, and Soonhee Han, ―Fish Disease Diagnosis System Based on Image Processing of Pathogens' Microscopic Images,‖ 2007 Frontiers in the Convergence of Bioscience and Information Technologies, 11-13 October. 2007.

[13] Sulistyo, S.B.; Wu, D.; Woo, W.L.; Dlay, S.S.; Gao, B. Computational Deep Intelligence Vision Sensing for Nutrient Content Estimation in Agricultural Automation. IEEE Trans. Autom. Sci. Eng. 2018, 15, 1243–1257. [Google Scholar] [CrossRef]

[14] Long Nguyen, Dongyun Lin, Zhiping Lin and Jiuwen Cao,― Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation‖, 2018 IEEE International Symposium on Circuits and Systems (ISCAS), May 2018.

[15] The Architecture and Implementation of VGG-16, available at << https://medium.com/towardsartificial-intelligence/the-architecture-and-implementation-of-vgg-16-b050e5a5920b >>, last accessed on 05 October, 2021 at 12:32 AM.

[16] 1. Dyrmann, M.; Karstoft, H.; Midtiby, H.S. Plant species classification using deepconvolutional neural network. Biosyst. Eng. 2016, 151, 72–80. [Google Scholar] [CrossRef]

[17] Yiming Ding, Jae Ho Sohn, Michael G. Kawczynski, ―A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain ‖, RSNA. Published online, Vol. 290, No. 2,pp. 456 – 464, Nov 06 2018 .

[18] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov and Liang-Chieh Chen , ―MobileNetV2: Inverted Residuals and Linear Bottlenecks‖, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, DOI: 10.1109/CVPR.2018.00474.

[19] MobileNet version 2, available at<< https://machinethink.net/blog/mobilenet-v2/ >> , last accessed on 05 October 2021 at 02:50 PM

[20] Franc¸ois Chollet ,―Xception: Deep Learning with Depthwise Separable Convolutions‖, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),July 2017.

[21] Review: Xception — With Depthwise Separable Convolution, Better Than Inception-v3 (Image Classification), available at << https://towardsdatascience.com/review-xception-with-depthwise-separableconvolution-better-than-inception-v3-image-dc967dd42568 >>, last accessed on 05 October, 2021 at 5:40 PM

[22] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object detection with deep learning: A review," IEEE Transactions on Neural Networks and Learning Systems, pp. 1–21, 2019.

[23] S. Zhang, X. Wu, Z. You, and L. Zhang, "Leaf image based cucumber disease recognition using sparse representation classification," Computers and Electronics in Agriculture, vol. 134, no. March, pp. 135–141, 2017.

[24] S. P. Mohanty, D. P. Hughes, and M. Salathe,´ "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, pp. 1–10, 2016.

[25] G. Geetharamani and A. Pandian, "Identification of plant leaf diseases using a nine-layer deep Convolutional Neural Network," Computers & Electrical Engineering, vol. 76, pp. 323–338, 2019.

[26] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, Feb. 04– 09, 2017.

[27] B. C. Karmokar, M. S. Ullah, M. K. Siddiquee, and K. M. R. Alam, "Tea leaf diseases recognition using neural network ensemble," International Journal of Computer Applications, vol. 114, no. 17,pp. 27–30, 2015.

Detection of Disease in Tea Leaves Using Machine Learning Approach

ORIGINALITY REPORT

20%	19%	8%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	12%
2	Submitted to Daffodil International University Student Paper	2%
3	Submitted to Liverpool John Moores University Student Paper	1%
4	www.mdpi.com Internet Source	1%
5	www.coursehero.com Internet Source	1%
6	Submitted to University of Lincoln Student Paper	<1%
7	Devi E M Roopa, R Rajadevi, R Shanthakumari, E Praveen, S SethuRaj, A C Shyam. "Mortality Prediction of Lung Cancer from CT Images Using Deep Learning Techniques", 2022 Fifth International Conference on Computational	<1%