PLANT DISEASE DETECTION BY USING TENSORFLOW

BY

NOWRIN AKTER
ID: 162-15-7811

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

Supervised By

Ms. Nazmun Nessa Moon
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2021

APPROVAL

This Project titled **"Plant Disease Detection by using TensorFlow**", submitted by Nowrin Akter, ID:162-15-7811 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 **JANUARY 2021.**

BOARD OF EXAMINERS

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

Internal Examiner

Gazi Zahirul Islam
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Raja Tariqul Hasan Tusher
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Dewan Md. Farid
Associate Professor
Department of Computer Science and Engineering
United International University

DECLARATION

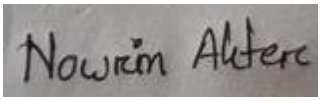
I hereby declare that, this project has been done by us under the supervision of **Ms. Nazmun Nessa Moon, Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

_____

Ms. Nazmun Nessa Moon
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



_____

Nowrin Akter
ID:162-15-7811
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project successfully.

I really grateful and wish my profound my indebtedness to Ms. **Nazmun Nessa Moon, Assistant Professor,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of AI to carry out this project. Her 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.

I would like to express my heartiest gratitude to the Almighty Allah and Head, Department of CSE, for his kind help to finish my project and also to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank my entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

My research paper titled "**Plant Disease Detection by using TensorFlow**'' which focus on our farmer's basic problem. Bangladesh is an agricultural country. Every year farmers of Bangladesh produce huge amounts of crops and we fully depends on it. But it is not so easy to take care of crops. There are many diseases that can harmfully effect our crop. Most of the time we cannot identify those diseases and also cannot take proper step to solve our problems. So I think about this and try to do something. Now everything became smart and technology dependent. We can see that the use of Artificial intelligence (AI) and machine learning everywhere. Following this, I built here a system. A system that is smart, fast and that system can identify the diseases of our crops. I use here python CNN model and TensorFlow library to classifying the diseases of crops. This system can make our work easy with big farm to identify the disease of tree or crops. We hope that it will make our work more accurate in less time and farmer of Bangladesh will also get benefits from here.

TABLE OF CONTENTS

CHAPTER 3: RESEARCH METHODOLOGY

CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION

CHAPTER 5: SUMMARY,CONCLUSION,RECOMMENDATION AND IMPLICATIONS FOR FUTURE RESEARCH

LIST OF FIGURES

LIST OF TABLES

TABLES NAMES **PAGE NO.**

CHAPTER 1

INTRODUCTION

1.1 Introduction

Image classification is a most important study material for exploring artificial intelligence (AI) and deep learning study. Image classification is one of the maximum used technologies and research topics to solve our problems of daily life issues and facts. Every time in our daily life we are facing many kinds of problems which can be easily handle and solved by using artificial intelligence (AI) and classification. Such a problem like knowing which disease effected our crops or plant. It is a common problem in our country that farmer cannot identify the main problem or main diseases of plants or their crops. So most of the time they cannot take proper step or proper solution to get good crops. I will use artificial intelligence and algorithms search a CNN and regression algorithms to find which disease effect our plants or crop and by identify the disease name we can take proper solution.

1.2 Motivation

Nowadays lots of farmers in our country suffer so much for the disease of their crop. Most of the time they cannot identify the main problem. They detect the disease by hearing from here and there and most of the time that is incorrect. Because of that they take wrong solution or medicine and that is dangerous. So I tried to give a proper solution to those farmer of our country. By scan

the leaf of tree or plant by mobile farmer can easily identify the disease name and then they can take perfects solutions. This way it will help them to grow good crops and more crops. So this research is focused to minimize this problem and creating a system with image processing which will make user able to scan fast and get the result.

1.3 Rationale of the Study

There is no doubt that there are many works at the Convolutional neural network (CNN) in image classifier sector. I can see many works that are done with images data and some of them can find cancer cell or some of them can detect covid-19. But working with the Google TensorFlow library is more effective and easier. It can show perfect output with less train data and this the matter I feel interested to work with this.

At the present time, I can see the use of artificial intelligence (AI) in every sector and we can also see the huge size of bad effect of plant disease in our agriculture. So I decide to do something new to detect the name or type of disease of plant by artificial intelligence (AI).

1.4 Research Questions

- Can I detect the disease name?
- Can I use it on our smart phone?
- Which kind of people will get the benefited?
- How much accuracy can we get?

1.5 Expected Output

Expected outcome of this research based project is to build a system to help farmer of our country and the people who need plant disease detector.

1.6 Report Layout

In **chapter 1,** all about this project is written here. The reason of choosing this project, how will this project be completed, project motivation, expected outcome and so on is discussed briefly. In a word, chapter 1 is the elaboration of introduction of my thesis project.

In **chapter 2**, related works on this area which were studied are showed. Their findings and limitations are summarized and hence the scope and challenges of the research are also mentioned.

In **chapter 3:** research methodology will discusses Research Subject and Instrumentation, Data Collection Procedure, Statistical Analysis and Implementation Requirements

In **chapter 4:** In here I discus about my Experimental Results and Descriptive Analysis

In **chapter 5:** presents a short conclusion. And list of reference

CHAPTER 2
BACKGROUND

2.1 Introduction

In this chapter, I will properly discuss some related research or project about image classification or related to detect something with image data. In the first part I will briefly discuss about some previous related work with image and in the second section I will show the outcome or a summary of my study of the associated work and then I will discuss about some good side and the challenge that I faced during this project work.

2.2 Related Works

In a research work that done by Deepika Jaswal, Sowmya, K. P. Soman [1] in Image Classification Using Convolutional Neural Networks classified different kinds of image, scene and environments. They used here CNN layers to make the classifier to do the classification. They proved that CNN can work properly with both face and non-face classification, building and nature classification, dense and rational classification. In their research work they used 250 epoch in their classification. They got 77% accuracy in forest and agricultural classification and about 91% of accuracy in residential vs agricultural classification. They got 97% of accuracy by working with green regions. They also proved that CNN gives proper and the best accuracy in the green scene rather than the building scene. So it works not only with material or building images but also with nature and leaf. So from this research we can ensure that CNN is a great algorithm to work with nature and nature products like leaf and leaf disease classification.

A research done by Bingquan Huo and Fengling Yin [2] on the topic Research on Novel Image Classification Algorithm based on Multi-Feature Extraction and Modified SVM Classifier. In their research work they worked with image object classifiers and SVM, Kernel and optimization. They used a common vector machine (SVM), K-neighbors neural networks and random forest to

make the model in order to classify images data. They find out that SVM has the best classification result or accuracy and it reduced amount of training time. They success to overcome the most common problem in computer vision classification and object detection by their model. But their accuracy and result accuracy were not very high.

So, I can decide from the research that CNN will work better than SVM in scene and nature image classification.

2.3 Comparative Analysis and Summary

After researching some research papers and project works I decided to work with CNN(Convolutional Neural Network) because,

- It works best with nature image or scene detection, green object detection and classification, forest and environment classification. Even it works good with the plant diseases which all I need.
- CNN has the best accuracy among others image classification algorithm with the accuracy of 90% or above with proper training and labelling data,
- It's easy to use and have a lot of resources to work further
- It also works best in comparing images
- It also can compare with the diseases and leaf.

As I choose to work CNN (Convolutional Neural Network) as my primary classification model. So I will working with CNN layers and deep learning algorithms to complete my project work. I used here TensorFlow and keras in backend to implement model and Adam optimizer. My main goal was using an big database and anaconda. But because of the huge amount use of CoLab and its popularity and easy, fast implementation way I decided to work with google CoLab. I used here CoLab and google own GPU in runtime to make the best out of it. As a result I had to use mount drive for database use using google drive.

My main goal is making a classifier of plant disease and further comparing them in order to get the output. So I used the Pooling layer, Con2d and activation layers for creating CNN layers. I worked with Apple, Corn and other easy to use leaf as in first. I cannot use hard to detect object classifiers which may lead us to bad accuracy and output. By using Corn, Apple and Grape leaf I will be able to get a good accuracy range in order to find out the best comparison possible among their range values. As if I have a good range between disease type then I will be able to say what is the disease name.

2.4 Scope of the Problem

The main focus of this research work or project is make a system that can identify the disease name of plant and proved that CNN is the best algorithm model for image classifier. In future I will try to use my system in big IOT based Agri farms.

I will make my system open source and freely available to farmer and all type o people. So that anyone like farmer or big farm owner can easily use my system and my data and get benefited from it.

2.5 Challenges

1) Data Collection

Since there are many available image data in online source, data collection was not so much hard for my research based project. But when I try start collect image data locally it become so much difficult for me because most of the farmer did not like that we touch theirs plants. For more accuracy and perfection I need more and more train data. But I was not able to take huge amount of plant leaf picture. Then I decide to collect data from **online.**

2) Model selection

Everyone know that model selection is a tough task for any researcher. Because research project success implicit on our Dataset and Model selection. The right decision will lead you at your goal easily and the wrong one will make it worst. I test various types of model with my test data and try to find out the best model that suit for my research work. I tried MATLAB too. After trying many way I decided to work with CNN algorithm because Google already did something great and big to make our work easy and faster. The moment I see that The Google built a library like TensorFlow for this type of images classifier research work and I can work with Google CoLab when working with images need an expensive type of GPU, but Google provides free virtual GPU for my research. So I decided to work on this model.

3) Data labeling

Data labeling was one of the important part of my research. Because it made code more faster. I used here **Bulk Rename Utility** to labeling my image data.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In in this section, I will describe my research methodology (Fig:3.1) and procedures. Moreover tools that I used for the research project, data collection, research topic, pre-processing, processing, statistical analysis and its implementation will be discusses in this session.

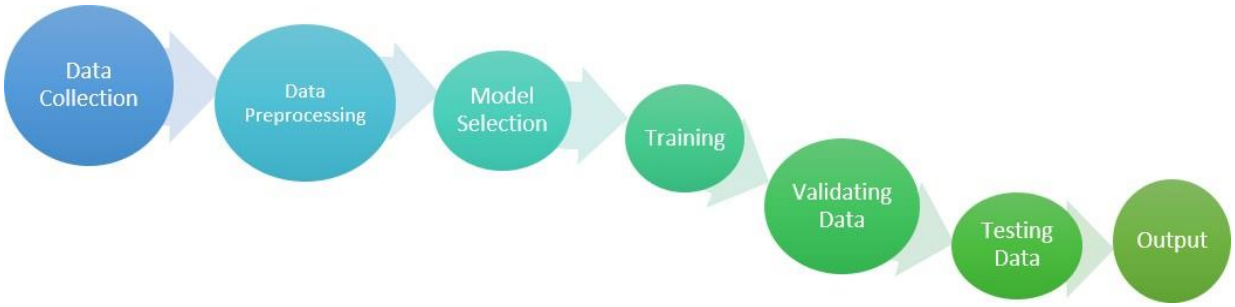


Fig 3.1: Methodology at a Glance

3.2 Research Subject and Instrumentation

I can see that data is the main part of the research. It is a very critical part for a researcher to find out proper data and best algorithm or model for his research. He/She also need to study hard about related research papers or projects. Then he/She should need to make some decisions:

- Which types of data should I collect?
- How to ensure that collected data are ready to use?
- How should each data be arranged?
- How should each data be named or labeled?

3.3 Data Collection Procedure

I used cats vs dogs image dataset as my experiment. I mainly selected this because those were easily available date which are already labeled and for my project I need labeled data. I used their 4000 cats and dogs images as training data and 2000 images as testing data

1) Data Pre-processing

Data pre-processing refers to the primary phase of processing datasets. Generally The raw datasets are not ready to perform operations and generate the outcome that we wand. As a result, data pre-processing is required. And it is considered to be one of the most important part of research. So, I try to collect more than 54305 image data from different way and from different sources and try to remove unnecessary or noisy data.

2) Data Organizing

In this phase I organized those data and store them in two data folder Val and train. Then I divided those test and train folder's data in 38 folder and those are Apple Scab, Apple Black Rot, Apple Cedar Rust, Apple healthy, Blueberry healthy, Cherry healthy, Cherry Powdery Mildew, Corn Gray Leaf Spot, Corn Common Rust, Corn healthy, Corn Northern Leaf Blight, Grape Black Rot, Grape Black Measles, Grape Leaf Blight etc.

3) Labeling Data

In this phase I renamed all of the image data as their name and also numbered them sequentially. I used here a software to labeling and the name of the software is 'Bulk Rename Utility' (fig:3.2).

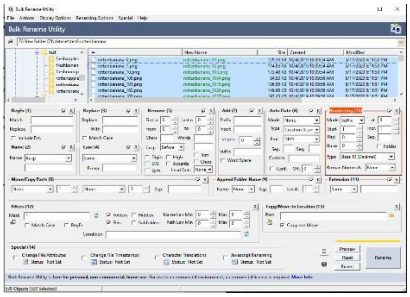


Fig 3.2: Data labeling by Bulk Rename Utility

4) Data Storing

In this part I move all of the file of data in google drive because it make our work easier in google drive. So, I can use those online stored data in my project by following some easy step or code.

3.4 Statistical Analysis

The amount of my total image data that I collect is more than 60,000 that I collect but after preprocessing I get total 54,305 data. Some data amounts that I use in my project (Table:3.1 and Table 3.2).

Table 3.1: Train image data amount

Fruits Name	Amount
Apple Black rot	693
Apple Cedar apple rust	581
Corn(maize) healthy	466
Corn(maize) Common rust_	442
Tomato Bacterial spot	524
Tomato healthy	595

Table 3.2: Test image data amount

Fruits Name	Amount
Apple Black rot	125
Apple Cedar apple rust	11
Corn(maize) healthy	138
Corn (maize) Common rust_	121
Tomato Bacterial spot	110
Tomato healthy	133

In here, I did not give all the actual number of my data. I just give this table to show the differences between the validation or testing data amount and training data amount.

3.5 Implementation Requirements

- **Python 3.8**

Python 3.8 is a version of Python programming language. It is a high level programming language. Most of the researcher used it to do their research work. It is highly recommended programing language for artificial intelligence (AI) based work or project and it is very popular among new generation's learner or programmer because it is very easy to understand and easy to implement.

- **Google CoLab**

  Google CoLab is a free and easy to use online base distributor of Python programming language. We can work here online by any kind of browser as like as we work in Jupiter notebook , but the main benefit of this Google CoLab for the reason that I always recommend to everyone is we can use it any kind of pc and it provide us free online virtual GPU access .

- **Bulk Rename Utility**

  It is a useful software to labeling any kind of data as you wanted. It is free and easy to use.

- **Hardware/Software Requirements**
  - Operating System (Windows 7 or above)
  - Web Browser(preferably chrome)
  - Hard Disk (minimum 4 GB)
  - Ram(more than 4 GB)

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

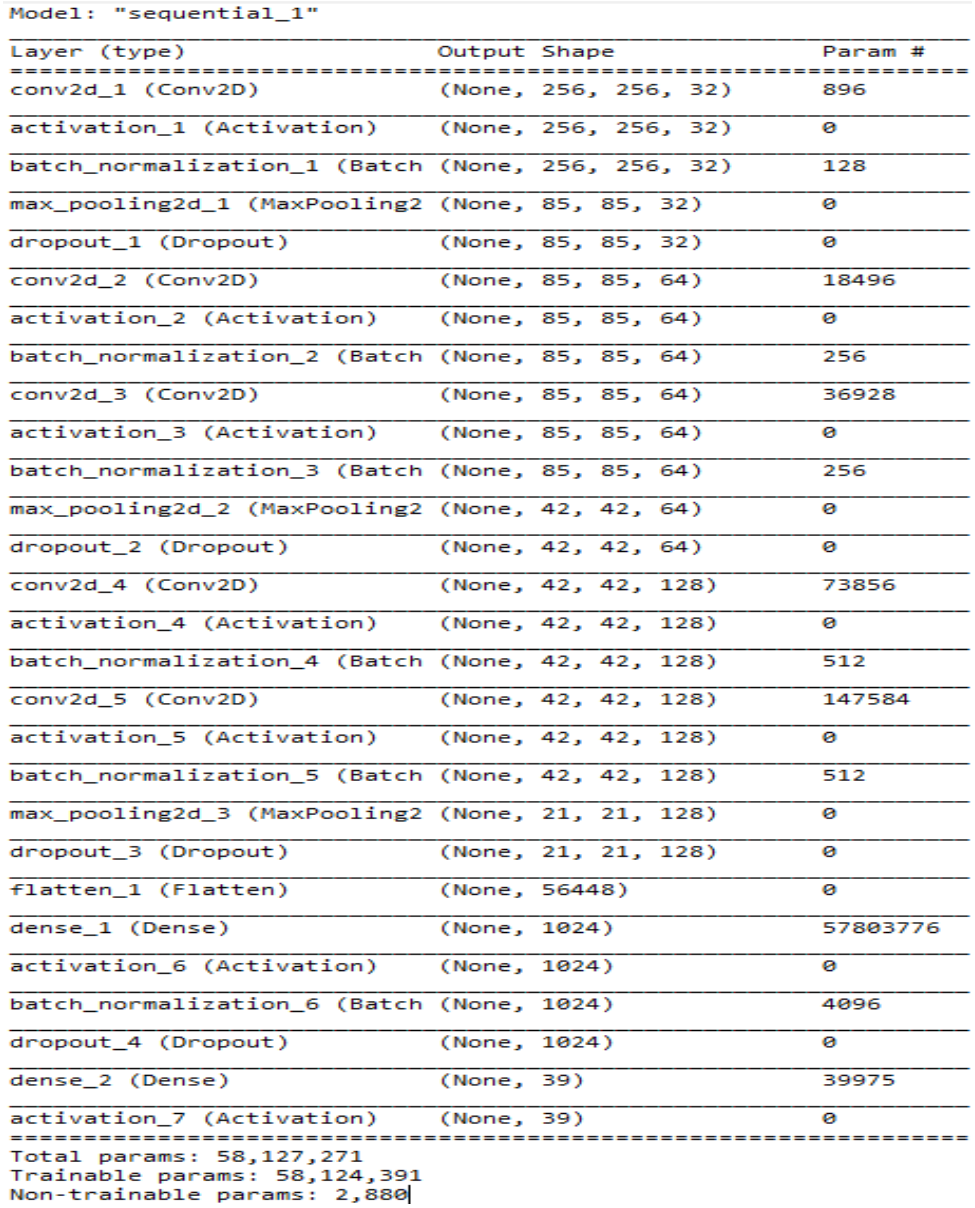
4.1 Experimental Setup

In model implementation and code implementation I collected data first and the procedure given below,

- As I worked with diseases detection I had to collect both good and bad plant leaf images from different areas
- For a larger part of my project I have collected images from the online.
- I also collect data from my garden.
- After that I finalized and normalized the data in order to perform the training.
- After labelling the data it was usable and good for further processes.
    - Then I preprocess my image in those steps,
    - At first, I crop images with same dimensions
    - After that I resize and reshape all the images in order to normalize dataset
    - Then I use a pooling layer to remove all of the negative pixels and make final usable images.

4.2. Model Summary

In my project model I used Adam optimizer an a sequential model with layers of conv2d, maxPooling2 then finally a flatten layer. There are five CNN layers (fig:4.1). Finally I wrap these layers with a flatter layer and a dense layer. As a result we get the accuracy with my model. From other research I found that more than three layers is best for creating a CNN model. By adding a dense and a flatten layer we can optimize the model and get proper output from the model with best accuracy.



```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 256, 256, 32)      896
_________________________________________________________________
activation_1 (Activation)    (None, 256, 256, 32)      0
_________________________________________________________________
batch_normalization_1 (Batch (None, 256, 256, 32)      128
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 85, 85, 32)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 85, 85, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 85, 85, 64)        18496
_________________________________________________________________
activation_2 (Activation)    (None, 85, 85, 64)        0
_________________________________________________________________
batch_normalization_2 (Batch (None, 85, 85, 64)        256
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 85, 85, 64)        36928
_________________________________________________________________
activation_3 (Activation)    (None, 85, 85, 64)        0
_________________________________________________________________
batch_normalization_3 (Batch (None, 85, 85, 64)        256
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 42, 42, 64)        0
_________________________________________________________________
dropout_2 (Dropout)          (None, 42, 42, 64)        0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 42, 42, 128)       73856
_________________________________________________________________
activation_4 (Activation)    (None, 42, 42, 128)       0
_________________________________________________________________
batch_normalization_4 (Batch (None, 42, 42, 128)       512
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 42, 42, 128)       147584
_________________________________________________________________
activation_5 (Activation)    (None, 42, 42, 128)       0
_________________________________________________________________
batch_normalization_5 (Batch (None, 42, 42, 128)       512
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 21, 21, 128)       0
_________________________________________________________________
dropout_3 (Dropout)          (None, 21, 21, 128)       0
_________________________________________________________________
flatten_1 (Flatten)          (None, 56448)             0
_________________________________________________________________
dense_1 (Dense)              (None, 1024)              57803776
_________________________________________________________________
activation_6 (Activation)    (None, 1024)              0
_________________________________________________________________
batch_normalization_6 (Batch (None, 1024)              4096
_________________________________________________________________
dropout_4 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_2 (Dense)              (None, 39)                39975
_________________________________________________________________
activation_7 (Activation)    (None, 39)                0
=================================================================
Total params: 58,127,271
Trainable params: 58,124,391
Non-trainable params: 2,880
```

Fig 4.1: The CNN model used in my project

4.3 Experimental Results & Analysis

After I run my dataset and created the model I find out the output that I wanted. When I compared same type of leaf is 93% to 100 %, same type of life with defected leaf 89% to 90 and different type of leaf 49% to 70% table 4.1.

Table 4.1: Fresh vs. rotten accuracy

Same	Different	Same but good vs. defected
93%-100%	49%-70%	89%-90%

So, from the table it has cleared that we get an output comparison and we can say from that it is defected or not and can say the name of the disease.

We can also understand the differences between Training accuracy vs. validation accuracy
And Training loss vs. validation loss by graph (fig: 4.2).

Fig 4.2: Training accuracy vs. validation accuracy
And Training loss vs. validation loss

4.4 Discussion

After studying my model and dataset I can say that this research has been concluded that classifier can be used in any kind of comparing dataset to compare and predict the accuracy among them. I successfully defined the accuracy of 93% of predicting the same type of good or defected leaf. It is also able to differentiate between diseases. So, we will be able to predict the name of disease.

CHAPTER 5

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

5.1 Summary of the Study

This research project is developed for find out the leaf disease and prove that CNN is the best image classifier algorithm. For this research I need some data which were image data. So those datasets were collected from the different online sources. After that data pre-processing rules maintained all the data and make them compatible with the systems environment. Datasets are trained for data handling purpose.

5.2 Conclusions

In this era of technology make new opportunity as well as new complexity. To handling new data it required new method, sometimes new technology. Finding right disease name that defected our plant is too hard for us. Sometimes we give wrong treatment because we cannot identify the real disease name. So this simple system can solve those problems. This is a simple system but this system is efficient.

5.3 Recommendations

Many algorithms are develop to identify objects from image data. Because our future is artificial intelligence (AI) based and image data in machine learning is very important for AI sector. It make our technology more effective and more productive. Everyone should need to work with image classification. It will change the full concept what we have been understood earlier.

5.4 Implication for Further Study

- To make this more efficient I shall collect more data
- I can use more class or type of data .
- In future I will build a complete open source working platform with huge amounts of image data.

REFERENCES

[1] Deepika Jaswal, Sowmya.V, K.P.Soman – ”Image Classification Using Convolutional Neural Networks”,-2014

[2] Bingquan Huo and Fengling Yin-” Research on Novel Image Classification Algorithm based on Multi-Feature Extraction and Modified SVM Classifier”, -2015

[3] Samer Hijazi, Rishi Kumar, and Chris Rowen, IP Group, Cadence, -” Using Convolutional Neural Networks for Image Recognition”.

[4] An introduction to convolutional neural networks [Online]available at:

[5] http://white.stanford.edu/teach/index.php/An_Introduction_to_Convolutional_Neural_Networks

[6] Tutorial on deep learning [Online] available at: http://deeplearning.net/tutorial/lenet.html

[7] Database of plant leaf, https://www.kaggle.com/, Last accessed on 2nd September, 2020

APPENDICES RESEARCH REFLECTION

During project activities, I faced many problems. But two problems ware major among them is selecting the algorithm and collecting data. Before working with CNN algorithm I tried more than three way to solve my problem and I failed to get perfect output. To collecting data I faced too much difficulty. Because locally data collection is time consuming and hard for student in this covid-19 pandemic time. So, I have to collect all of my data from online sources. And after a long time and a lot of attempts and hard work I got succeed.