Machine Learning Approach to Predict Tomato Leaf Disease

BY
SHUMAYA AKTER
ID: 221-25-138

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

Supervised By

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

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

5th August 2023

APPROVAL

This Project/Thesis titled **Machine Learning Approach to Predict Tomato Leaf Disease,** submitted by Shumaya Akter, ID No: 221-25-138 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 M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 05-08-2023.

BOARD OF EXAMINERS

Chairman

Dr. Sheak Rashed Haider Noori, PhD
Professor and Head (In-Charge)
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md. Atiqur Rahman, Phd
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Mr. Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

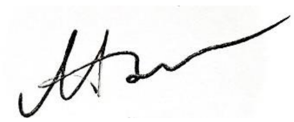
External Examiner

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

DECLARATION

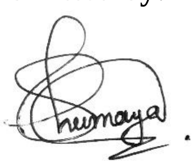
I hereby declare that, this thesis has been done by me under the supervision of **Ms. Nazmun Nessa Moon, Associate Professor, and Department of CSE** Daffodil International University. I also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree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

Supervised by:



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

Submitted by:



Shumaya Akter
Student ID: 221-25-138
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

At first, I express my heartiest thanks and gratefulness to Almighty Allah for His divine blessing makes me possible to complete the final year thesis successfully.

I really grateful and wish my profound my indebtedness to Ms. **Nazmun Nessa Moon, Associate Professor,** Department of CSE Daffodil International University, Dhaka. Machine Learning interest of our supervisor in the field of AI to carry out this thesis. 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 thesis.

I would like to express my heartiest gratitude to the Almighty Allah and thanks to **Dr. Touhid Bhuiyan, Professor and Head, Department of CSE**, for his kind help to finish our thesis 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 passion of our parents.

ABSTRACT

The machine learning approach used in this thesis uses a large dataset of leaf photos and associated disease labels to predict tomato leaf diseases. A large number of tomato leaf photos, comprising both healthy and diseased leaves, is gathered for the research. Techniques for preprocessing are used to enhance picture quality and extract relevant characteristics. The objective of the technique is to find distinguishing traits in leaf pictures that distinguish between various disease classifications. Convolutional neural networks (CNNs), decision trees, random forests, support vector machine, and other machine learning methods are assessed for disease prediction. To evaluate the performance of the models, the cross-validation methods are used to verify them on the labelled dataset. To learn more about the visual patterns connected to each condition, feature importance analysis is done. Techniques for transfer learning are investigated to make better use of taught models. The experimental findings show a high degree of prediction accuracy for tomato leaf diseases, with transfer learning-based CNN-based models outperforming more conventional methods. The research helps create an automated method for early disease identification, which helps farmers and specialists execute effective disease control techniques on time, reducing crop losses, and maintaining sustainable tomato production. The paper highlights the potential of machine learning in plant pathology and encourages further investigation into related methodologies for additional plant species and disease classes.

TABLE OF CONTENTS

CONTENTS **PAGE NO**

LIST OF FIGURES

LIST OF TABLES

TABLES **PAGE NO**

CHAPTER 1
INTRODUCTION

1.1 Introduction

To support the anticipated demand, worldwide harvest production needs to increase by at least half by 2050. Currently, the majority of creation takes place in Africa and Asia, where 83% of ranchers are family businesses with little or no agricultural experience. Along these lines, yield hardships of more significant than half; because diseases and vermin are common. At this time, it is not possible to characterize crop diseases using the traditional method of human examination through visual investigation. The development of PC vision models provides a quick, uniform, and precise solution to this problem. A classifier can also be distributed as an application once it has been prepared. Easy to utilize, all things required is a web affiliation and camera prepared wireless. Standard business applications I-Naturalist and Plant Snap display how this can be executed. The two applications have achieved accomplishment in not simply passing authority on to clients yet furthermore in building a natural on the web informal organization. The availability and cost of cell phones continue to rise annually. There will be approximately 5 billion cell phone users worldwide in 2020 [1]. Of this, one billion clients are arranged in India and a further one billion are tracked down in Africa. According to Statista, these numbers have consistently increased annually over the past ten years. With these real factors as a fundamental need, it is acknowledged that simulated intelligence applications will play a critical work in trim the possible destiny of developing. Such headways have helped with diminishing planning time and mix-up rate. Most importantly, the development of engineering has been a major focus of the vast and complex datasets of the 21st century [3]. one design in progress; Res-Net offered additional significant capabilities. This combines substantial group standardization with dynamic skip associations. This makes it possible to prepare at a much faster learning rate. Wu et al., in 2019, compared Res-Net to VGG-Net, Google-Le-Net, and Dense-Net, and came to the conclusion that Res-Net produced the best results when ordering grape leaf diseases. In current investigation; structures including Alex-Net, Le-Net and Google-Net, are conventionally intertwined into the groundwork of custom structures. Wall-Elgin came up with this idea; in light of Le-Net's investigation of the order of soybean diseases. The model involved three convolution layers, one max-pooling layer and a totally connected MLP with Re-Lu incitation and achieved a 91% exactness rate [1].

The preparation of information prior to the model's execution is extremely important. Viral, bacterial and infectious illnesses can be difficult to perceive, consistently sharing a front of signs. These signs can be any measurable change in color, shape, or capacity as the plant reacts to the microbe. Considering this diverse nature, it is advantageous over use RGB data. This produces clear, upheaval free pictures which might take more time than greyscale data to get ready, yet by and large more fitting for plant infection recognizing confirmation models.

A model's dependability can be affected by smaller datasets or inconsistencies in information. This can be monitored in multiple ways using methods like growth or move learning. In addition to reducing overfitting, enlarging preparation images has the potential to enhance a model's overall execution. This can be accomplished by adding features like zoom, pivot, and color or differentiation changes. The changed pictures ought to, regardless, reflect the suspicions of the endorsement dataset. Even with the addition of new data, a classifier's accuracy can suffer when used incorrectly.

When working with smaller datasets, the move learning strategy has also proven to be very effective. This incorporates changing the heaps of a pre-arranged model. The ImageNet data base is ordinarily used hence and contains north of 14 million pictures.

Mohan-ty et al. in 2016 uncovered these advantages in an assessment focused in on gather illness game plan. Here, unmatched results were recorded using move learning ImageNet, diverged from a model worked without any planning. Since ImageNet contains pictures that aren't needed for a plant-specific assignment, it's not clear if pre-preparing on a herbal information base can improve performance. According to ebb and flow research, pre-preparing on ImageNet may result in better results, but pre-preparing on a plant-specific assignment may reduce overfitting. Regardless, these assertions are speculative. The topic is typically ignored because large organic datasets do not exist. Growth can also be used with relevant models. The abilities of the model are greatly influenced by the quality and type of prepared information. Exactly when ready on imagery which contains plain establishment data, a classifier's precision gets subject to this creation.

By separating a leaf from its experience, segmentation for this situation may prove successful. This methodology can other than be used in conditions where the classifier requires scene care. This could mean knowing how much damage is being caused by microbes all around the contaminated tissue as opposed to just the contaminated tissue. Segmentation is not just another concept; it has been used to group diseases since the 1990s. Indeed, excellent outcomes were accounted for even at this early stage. Additionally, early tests were helpful in identifying the limitations, demonstrating that the strategy could not compete with helpless picture quality. Consequently, putting an emphasis on the significance of careful information collection and preprocessing. Segmentation's relevance continues into 2020 [1]. Combining this with particular symbolism has a lot of potential for investigation.

The stage of the disease and possible location are also determined by the type of preparation information used. For early infection acknowledgment, unequivocal imagery ought to be used. Chlorophyll fluorescent (CFI), infrared thermography (IRT), hyperspectral (HSI) and multispectral (MSI) imagery have express abilities to recognize appearances which are not yet perceptible to the independent eye. These can be used on their own or combined when necessary.

For model, IRT has the noteworthy ability to perceive a development in temperature. Due to the limited availability of such information, the topic of early location is somewhat neglected. With a rising scholastic premium in the region, the innovation anticipated to capture this particular symbolism is becoming more moderate. In any case, it is not even a tool that distant ranchers can use at this point. As needs be, it is crazy to recollect it for an endeavor expected for such clients.

1.2 Motivation

Tomato plants are among the most broadly developed and monetarily significant harvests around the world. However, they are highly susceptible to a variety of deadly diseases that have the potential to have a significant impact on their growth, productivity, and quality. Convenient and precise location of tomato leaf illnesses assumes an imperative part in forestalling the spread of contaminations and executing successful sickness the board methodologies. Conventional strategies for sickness distinguishing proof and conclusion frequently depend on manual assessment by specialists, which can be tedious, emotional, and inclined to blunders. In fields like healthcare, finance, and image recognition, machine learning has emerged as a potent tool. Its capacity to dissect enormous datasets and separate significant examples has made it a promising methodology in the field of plant pathology. Applying AI procedures to foresee tomato leaf infections can offer a few benefits over conventional strategies. It has the potential to significantly improve the accuracy and efficiency of disease detection by utilizing the power of automated analysis and pattern recognition. This would result in improved disease management and increased crop productivity. Besides, the rising accessibility of advanced imaging advancements and the developing interest in accuracy farming give a superb chance to foster PC vision-based approaches for sickness expectation in plants. Researchers can use the vast amount of information contained in leaf images to identify disease patterns that may not be readily apparent to the human eye by utilizing machine learning algorithms. This can empower early location and mediation, forestalling the spread of infections and limiting the financial misfortunes brought about by ranchers. Farmers and agricultural experts can make informed decisions about disease management strategies, optimize resource allocation, and reduce the need for excessive pesticide use by developing an automated system that can accurately predict diseases based on leaf images. Eventually, this proposition tries to advance economical tomato creation, improve food security, and prepare for the reception of AI strategies in plant illness expectation across various harvests and districts. Reasoning of the Review The event of tomato leaf sicknesses represents a huge danger to tomato crop efficiency and quality. Effective disease management and a reduction in farmers' economic losses require prompt detection and precise prediction of these diseases. Time-consuming and subjective, traditional methods of disease identification necessitate expert manual inspection. AI methods have shown extraordinary commitment in different spaces and might possibly alter sickness expectation in plants.

Nonetheless, there is a restricted group of examination explicitly zeroing in on the utilization of AI in foreseeing tomato leaf sicknesses. By developing a machine learning strategy for accurate disease prediction based on images of tomato leaves, this study seeks to fill this knowledge gap. By utilizing the force of computerized investigation and example acknowledgment, AI calculations can handle enormous datasets and extricate significant illness designs that may not be quickly detectable to the natural eye.

The results of this study will add to the logical comprehension of AI's part in plant pathology and give useful ramifications to the farming business. By fostering a computerized framework for tomato leaf sickness expectation, ranchers and horticultural specialists can settle on informed conclusions about illness the board methodologies, advance asset distribution, and lessen the dependence on substance therapies. At last, this study looks to improve tomato crop efficiency, diminish monetary misfortunes, and advance maintainable horticultural practices.

1.3 Rationale of the Study

The rationale for this thesis work, which uses machine learning to predict tomato leaf disease, comes from the acute need to deal with issues in contemporary agriculture. Because they are a valuable worldwide commodity, tomato crops are seriously threatened by a number of illnesses, which has an impact on farmers' yields and their bottom line. Traditional disease detection techniques often fall short on accuracy and efficacy, impeding timely responses and escalating the usage of chemical pesticides that might have an adverse impact on the environment. This work intends to create a prediction model that can quickly and precisely diagnose tomato leaf diseases using the power of machine learning algorithms, providing farmers with early diagnosis and precise disease control measures. This study might help increase agricultural output, support environmentally friendly practises, and promote food security. A persuasive argument for continuing this research is that investigating the use of cutting-edge technology in agriculture may provide the groundwork for further advancements in the field of disease prediction and control.

1.4 Expected Output

- Detect Disease with 90% above accuracy.
- A graph will show the effecting level of this research
- An output will give visualization of disease or healthy leaf.

1.5 Report Layout

Chapter 1: In this part, I may look at the beginning, Inspiration, Objectives, or Expected Outcome, among other things is rapidly examined. Simply, part 1 of this project elaborates on its presentation.

Chapter 2: In this part I had presented about the fundamental condition of our work and talk over rule in the endeavor. We also talk about the deals that are related to this area and were looked at, which are also shown. Their findings and limitations are summarized, and the examination's length and difficulties are also mentioned.

Chapter 3: In this section, I discussed the study methodology, including the assessment topic and instruments, data grouping, real examination, and execution requirements.

Chapter 4: I spoke about trial findings, outcomes from experiments, and descriptive statistics in this chapter.

Chapter 5: The effects of my research on society and the environment are discussed in this chapter.

Chapter 6: I detailed the whole process of this investigation, along with several sources, in this chapter.

CHAPTER 2

BACKGROUND

2.1 Introduction

Our lives are so quick and simple thanks to technology, where I want everything at our fingertips. I will discuss Deep Learning-related data classifier-related research in this chapter. The entire process of developing the model for tomato leaf disease recognition with extensive Inception 50, Resnet v3, and Resnet 152v2 is described in greater detail. I'll discuss earlier work that was connected to this one in the initial part. I will provide the findings, which are an overview of my research on the relevant topic, in the second part. Finally, I'll discuss the benefits & drawbacks of working on this job.

2.2 Related Works

Kim proposed the principal CNN model I portray underneath and a while later develop in this work. Zhang and Wallace [4] precisely investigated this model's properties. I similarly note that Zhang et al. extended this model to accommodate multiple arrangements of pre-prepared word embeddings together. Despite using one-hot vectors set up of (pre-prepared) word embeddings, Johnson and Zhang proposed a comparative CNN engineering nearly simultaneously to Kim [8].

Dhruva K. Bhattacharyya and A Chowdhury suggest a Co-Expression Analysis of Gene Expression: A Review of the Top Techniques Additionally, it gave some clarification on the analysis of scRNA-seq data and a summary of the best procedures for analyzing (differential) co-articulation, express systems, unequal networks administration, and unequal accessibility in RNA sequence data and microarrays. Ming hang, Yangquan Dai proposed Picture of wheeze infection division model ward on heartbeat coupled brain Organization with blend frog hop estimation [9].

For plant diseases, a half-and-half frog-bouncing-calculated epic picture segmentation model called SFLA-PCNN is proposed. Using the wellness capability of SFLA and picture segmentation minimization, the image of potato late scourge illness is employed as a preliminary segmentation picture to determine the appropriate setup borders of PCNN [4]. The process of extracting elements and identifying plant diseases from images has made significant progress thanks to picture segmentation. Chit Su proposed Plant Illnesses Acknowledgment for Shrewd Cultivating Utilizing b-Model based Measurable Elements. This demonstrated the benefits of employing the GP flow model for the Filter descriptor and has been used in the representation of plant illness [10]. The suggested highlight also strikes a good balance between execution and grouping precision. Despite the fact that our suggested feature can truly demonstrate the Filter's incorporation and use in plant sickness confirmation, it is necessary to make an effort to implement it by taking into account and collaborating with other image planning techniques.

2.3 Comparative Analysis and Summary

Humanity considers plants to be a source of power. Horticulture may be affected by tomato leaf diseases, which might have catastrophic effects on harvest yields. Therefore, identifying leaf diseases is essential for the agriculture industry. In any case, it requires colossal work, moreover dealing with time and wide data and aptitudes about leaf illnesses. AI then plays a role in identifying diseases in plant leaves by breaking down data from various sources and classifying it into one of a predetermined set of categories.

This investigation's precise objectives are to:

I. Utilizing both an approval dataset and a test dataset, determine the model's general viability in describing diseases.

II. Check the model's accuracy using a variety of picture sizes and settings for enlargement.

III. Send the pre-arranged model to make an easy to use an application.

The f1-score and exactness estimates will be looked at before the model is presented due to an uneven class dispersion. If the representation reaches an accuracy and F1score of greater than 80%, it will be acknowledged.

This inquiry is going to be completed taking micro agricultural requirements into account. Both the classifier and the web application need a phone and a web connection, both of which, as was already said, are becoming more widely available. The model will be tested with a variety of image sizes and development settings to understand the needs of basic camera phones. I have continued to exercise caution over excruciating accuracy and yield. I can locate the optimum link with amazing accuracy by applying estimates.

2.4 Scope of the Problem

I had to create the component such that anybody, regardless of programming knowledge, would be able to utilize it and get the necessary information. It suggested a system for anticipating leaf maladies. It offers explanation on the trial evaluation of our technique. For each condition, a different amount of images are collected and then categorized as informational and database images. The surface-arranged lighting and shape decide the fundamental credits of the image.

2.5 Challenges

a) Data Collection

A model's resolute quality might be impacted by using more basic datasets or undifferentiated data. This may be seen in a variety of ways, utilizing techniques like move learning or expansion, for instance. By enhancing image preparation, overfitting may be decreased and a model's overall performance can be enhanced. The kind and caliber of prepared information have a significant impact on the model's capabilities. A classifier's accuracy depends on this design when trained on imagery that comprises simple establishment data. Additionally, a vast amount of image data was supplied for greater accuracy. The need for such a dataset is strongly stressed throughout the test.

b) Model selection

The toughest and most important component of any research effort is undoubtedly model assurance. Any assessment's success depends on information gathering and model assurance. If I choose the right model, there should be a favorable result shortly. In fact, choosing the incorrect one can leave you disappointed. In order to find the most logical model for my research project, I have tested a few different ones using our test data. Anyhow, I came to the conclusion that the Naive Byes Classifier computation was the simplest and best choice for us based on all of my study. Once I learned that it offers a user-friendly library, I understood that the classifier job would be easier. I created this model as a result to be used in the project.

Image Pre-Processing and labeling

Information identification is also an essential part of the speedier code. It took up the most of our efforts. Since it makes running programmers quicker and data preparation less monotonous. The quality and arrangement of the photographs that were acquired from the Internet varied.

To ensure uniformity, the final images that would serve as the dataset for a deep brain network classifier were preprocessed, which improved highlight extraction. In order to emphasize the region of interest (the tomato leaves), the approach of picture preprocessing also included physically reducing the proportion of images to make a square around the leaves. A image with a modest purpose and an estimate less than 500 pixels was not regarded as a meaningful picture for the dataset during a social event. As a result, it was made sure that the photographs included all the information needed for feature learning.

Various resources can be found through looking across the Web, but their relevance is often volatile. Considering a real worry for confirming the exactness of classes in the dataset, at first accumulated by a watchwords search, horticultural experts examined leaf pictures and denoted every one of the photos with reasonable sickness truncation. Copy images that remained after the underlying emphasis of social event and grouping pictures into the classes depicted in Section 4.2 were removed from the dataset at this stage, the input data for this project shown in Figures 2.1, 2.2, and 2.3.

Fig 2.1: Example of pre-Processed Input Data.

Fig 2.2: An example of input data that has already been processed.

Fig 2.3: Example of phase O pre-Processed Data for Input Leaf

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The perception of selected, measured, and examined information is the research methodology. I'll discuss and illustrate our examination strategy in this part. Additionally, this conference will cover mechanical components for the investigation project, data gathering, the research topic, pre-planning, planning, authentic evaluation, and its implementation. I'll go into more depth about the study's methodology and processes in this part. On the other hand, this chapter will discuss the study's subject, data collecting, processing, and pre-processing, as well as the application of statistical analysis. The entire process is depicted in figure 3.1.

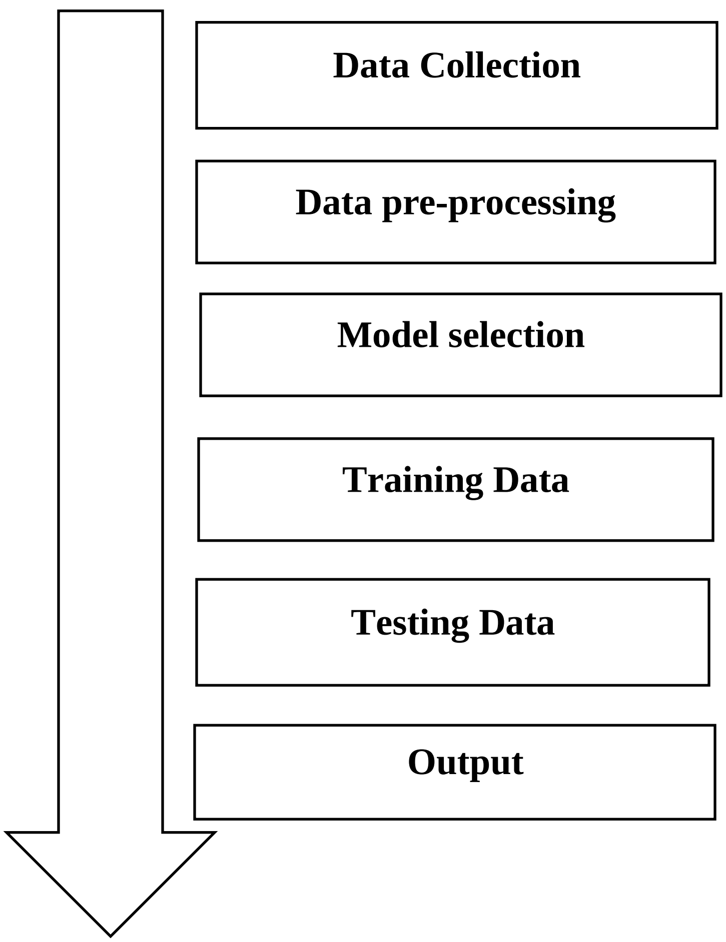


Figure 3.1: Methodology of this research

3.2 CNN – Plan of Attack

Dominating CNNs is undoubtedly not a simple feat. Since this is the first part of the excursion, you'll need to learn the basics first.

A sort of introduction to convolutional neural networks.

Step 1(a): Convolution Operation

The key tenet of our attack plan is convolution activity. In this step, I'll discuss indicators, which essentially serve as pathways for the brain organization. Additionally, I'll examine feature maps to understand about their limitations, how models are seen, the levels of recognizable evidence, and how disclosures are represented and exhibited given bellow in fig 3.2.

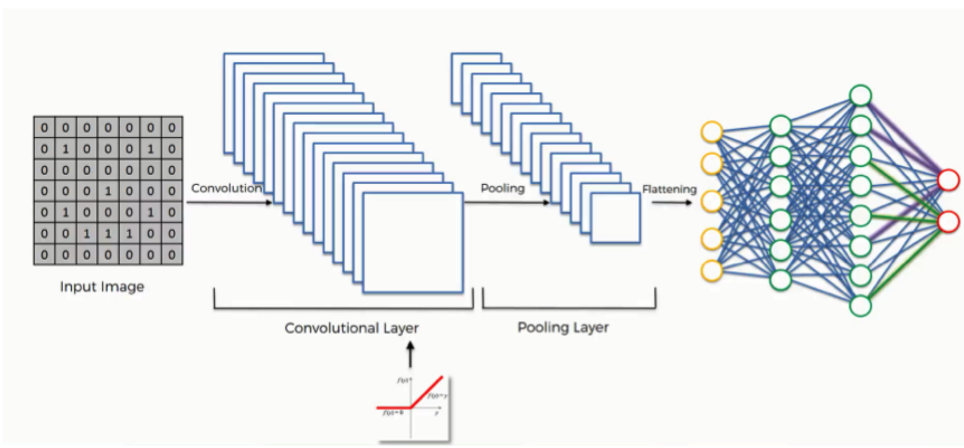


Fig 3.2: CNN's image classification steps

Step 1(b): ReLU Layer

The Rectified Linear Unit, or ReLU, will be the second component of this progression. Convolutional neural network linearity capacities will be examined in relation to the ReLU layers.

Rather than being redundant for comprehending CNN's, there is nothing wrong with vigorous exercise to enhance your abilities.

Stage 2: Pooling

I'll go over pooling in detail and see how it works in general in this section. In any case, there will be a particular kind of pooling at our nexus here; max pooling. I'll cover various philosophies, nonetheless, including mean (or aggregate) pooling. This part will end with a display made using a visual keen device that will sort the whole thought out for you.

Stage 3: Flattening

In this part, the rectification process and the change from pooled to levelled layers while using convolutional neural networks will be briefly addressed.

Stage 4: Full Connection

This part will summarize what I covered in the previous segment. You will have a better knowledge of how neural networks based on convolution operate and how the generated "neurons" gradually learn how to organize photos after you have learned this.

Synopsis: I'll finally wrap things up and provide a succinct overview of the idea covered in the section. If you believe it will help you (which it very likely will), you should check at the supplementary instructive activity that covers Soft-Max and Cross-Entropy. Although it is not required for the course, having a basic understanding of these concepts will come in handy when working with Convolutional Neural Networks. Delicate Max and Cross-Entropy: An extra to work on your cognizance of Convolutional Brain Organizations. Might it be said that you are down? So, let's get started!

3.3 Research Subject and Instrumentation

The most important component of the inquiry project is the data. Finding fantastic data and fantastic calculations or models for our examination work is a crucial component for experts. I similarly need to find out about related assessment papers. At that point, I must select one or more options:

1. Which data ought to be gathered?
2. That information that was gathered is okay?
3. How should all of the information be coordinated?
4. How should each piece of information be labeled?

3.4 Data Collection Procedure

I gather and combine a data bundle made up of thousands of tomato leaf photos. A model's resolute quality might be impacted by using more basic datasets or undifferentiated data. This may be seen in a variety of ways, utilizing techniques like move learning or expansion, for instance. By enhancing image preparation, overfitting may be decreased and an algorithm's performance as a whole can be enhanced. The kind and caliber of prepared information have a significant impact on the model's capabilities. A classifier's accuracy depends on this design when trained on imagery that comprises simple foundation data.

Fitting datasets are required at all periods of article affirmation research, starting from getting ready stage to evaluating the display of affirmation estimations. To perceive strong leaves from sick ones, another class was added in the dataset. It contains simply pictures of strong leaves. It was beneficial to include a foundation image class in a separate class in the dataset in order to get a more accurate order. The next stage was to increase the dataset's picture count.. The primary goal of this study is to get the company ready to learn about the characteristics that distinguish one class from others. As a result, the organization now has a better chance of understanding the relevant highlights when using images that are larger.

Data Pre-processing

Information and Data pre-planning is a taking care of that way to the pre time of dealing with datasets. All things considered rough data pictures can't perform undertakings and produce expected outcome. As a result, information pre-handling is necessary. On Fig. 2.1, I showed an example of a preprocessed input image.

Data Organizing

Data Images An arrangement of gathered data is information that is being sorted out. Along these lines, for straightening out data, I have attempted and arranged the data and saved them in two envelopes. In information arranging, I have also used approval organizer to verify train information approval.

Import Image Data

At this part we have some code in colab that import our information from google drive. And all picture information I need to transfer from the get-go in google drive. Some interaction is given underneath Fig 3.3.

```
[ ] training_set = train_datagen.flow_from_directory('/content/drive/MyDrive/Colab Notebooks/Data Set/Train data',
                                                     target_size = (224, 224),
                                                     batch_size = 16,
                                                     class_mode = 'categorical')

    Found 1539 images belonging to 10 classes.

[ ] test_set = test_datagen.flow_from_directory('/content/drive/MyDrive/Colab Notebooks/Data Set/Test data',
                                                target_size = (224, 224),
                                                batch_size = 16,
                                                class_mode = 'categorical')

    Found 510 images belonging to 10 classes.
```

Fig 3.3: Data importing using Colab python3.8

3.5 Data Processing Layer

In reality, each picture will undergo a sequence of convolution layers with channels (Kernels), pooling, totally associated layers (FC), and Soft-Max capabilities to arrange an item with probabilistic attributes ranging from 0 to 1 while deep learning CNN models are being developed and evaluated. The graphic below shows the procedure CNN employs to manage an information picture and arrange the articles according to their attributes.

3.6 Statistical Analysis

While I used a variety of algorithms, my primary objective was to increase accuracy. I got the required precision by means of CNN, Origin V3, RESNET50, RESNET152V2, VGG19, and Versatile NET. The process flowchart for my entire work is given in fig 3.4 :

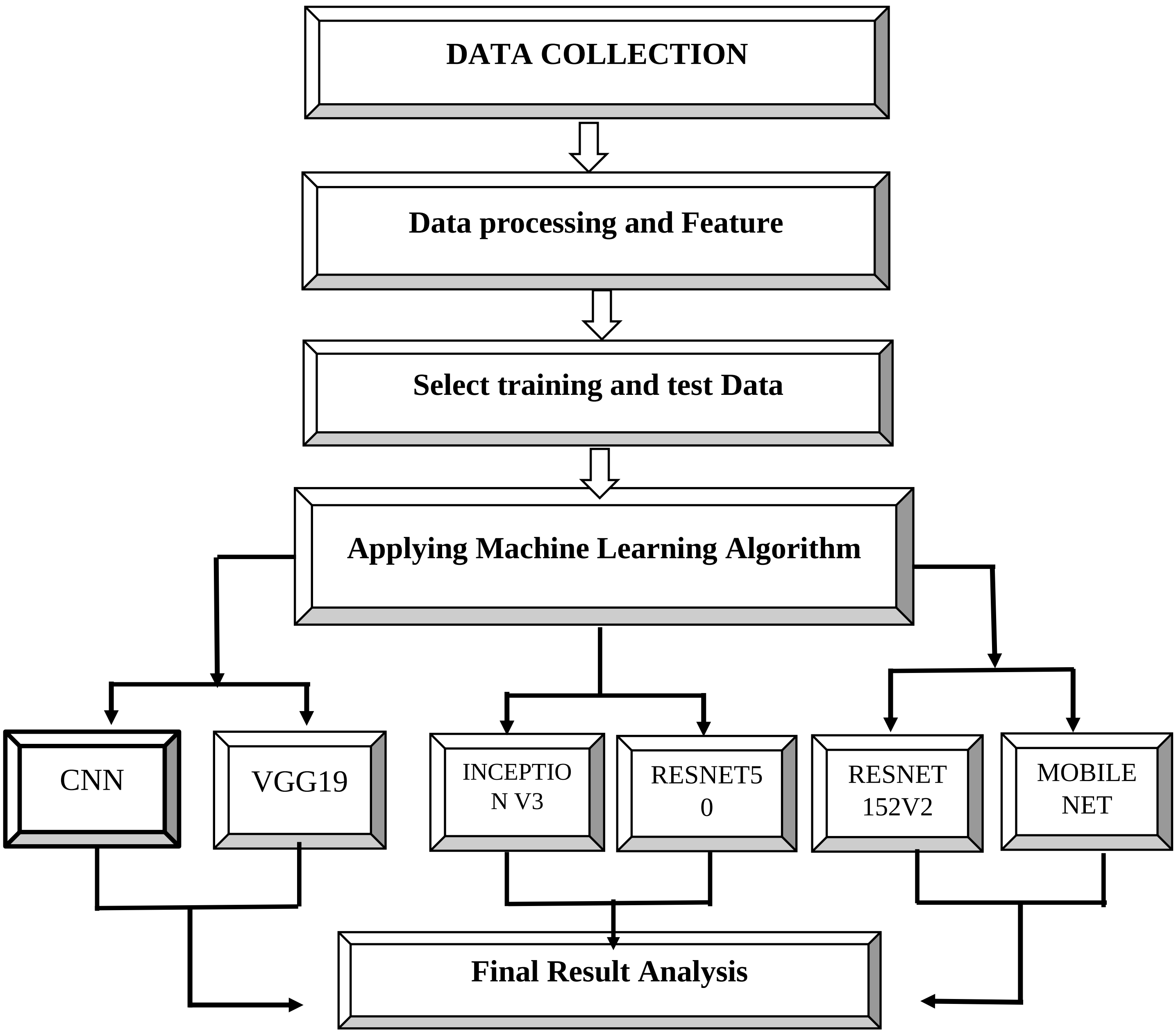


Figure 3. 4: Proposed Model Structure

I detail each step of my working procedures in this figure 3.4. My working flow has been depicted on the flow chart.

3.7 Implementation Requirements

- **Python 3.10.2**

  The most recent Python rewrite is 3.10.2. Python is a high-level programming language that is open-source, user-friendly, and hailed as the strongest computer language. Due to CNN, VGG19, InceptionV3, Portable Net, Resnet152v2, and Resnet50's very complex compositional schemes, using another language will be exceedingly challenging. Because they are so straightforward, Python's built-in functions and commands are easier to implement. Other programming languages that would have been used would have required a substantial commitment of time.

- **Google Colab/Jupyter Notebook**

  I can use my browsers to access cloud resources using Colab, a hosted Jupiter that has been installed and configured so that we don't need to utilize our PCs. It works very much like Jupiter. They are based on notebooks, which can be text, image, or code, as only the Python kernel can now be used instead of Jupiter Collab.

  A Jupyter notebook is also used to accomplish the same thing. Equipment abilities put them aside from each other. If you have a high-end build and an external GPU, the Jupyter notebook is ideal. Since it was cooperative work, I needed to utilize the two stages.

- **Hardware/Software Requirements**

  - Operating System (Windows 7 or higher or Linux)
  - Web Browser (Recent versions Chrome, Firefox, or Microsoft Edge)
  - Hard Disk (At least 120GB)
  - Ram (4 GB or more)
  - GPU (Minimum 2GB)

CHAPTER 04

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

On The following section shows the methods used to create and submit the classifier. Three steps, each of which focused on a distinct job, make up the CNN order process. All of the work in this investigation was done on a single machine. The hardware and software specifications are as follows: 8.0 GB of memory; 2.40 GHz Intel(R) CoreTM i5-9300H CPU; 6 GB of GDDR6 graphics from NVIDIA; Windows 10 Home 64 operating system. The initial step was to gather datasets for the development of my model and code. The variety of datasets required significant expenditure. Additionally, they were converted to jpg format.

4.2. Information Pre-Processing

20% of the dataset is used for approval, while 80% is used for preparation. Increase settings are first applied to the data that is being prepared. These decisions are made "on the fly," and each activity's likelihood of occurring at each age varies. Flipping (inconsistent), padding mode (reflection), and zoom with crop (scale = (1.0, 1.5)) are the parameters that have been implemented. After neglecting "Zoom with crop," it was discovered that it had improperly controlled leaf degeneration zones. All of the photos are then uniformly scaled. A pack operation is used to complete the resizing to 256 × 256. Since a pre-made model is employed, RBG ImageNet metrics are used to standardize. An illustration of the last pre-taken care of pictures is detectable in Fig 2.1

4.3. Arrangement by CNN

1) Testing image Size in Phase One: The first phase's goal is to determine how image size affects model execution. Five photos, ranging in size from 256 × 256 to 150 x 150, are said to have been tested collectively. As a starting point, the downloaded Resnet34 loads are utilised. In move learning, all layers—aside from the last two—are frozen by default. These are linked to the plant sickness request job and include new loads.

Thanks to freezing, such layers may be created without the back propagation of angles. The one-cycle approach is employed specifically in this way to prepare the final layers. The remaining layers with this total are delivered. To aid in the adjustment cycle, a plot that contrasts learning rate and bad luck is made. The model is then run using a logical learning that is selected from this. Every preliminary maintains predictability across all variables, including learning rate. What's more, the convolutional Activity handling in the Fig 4.1. This data from CNN Calcification Model [5].

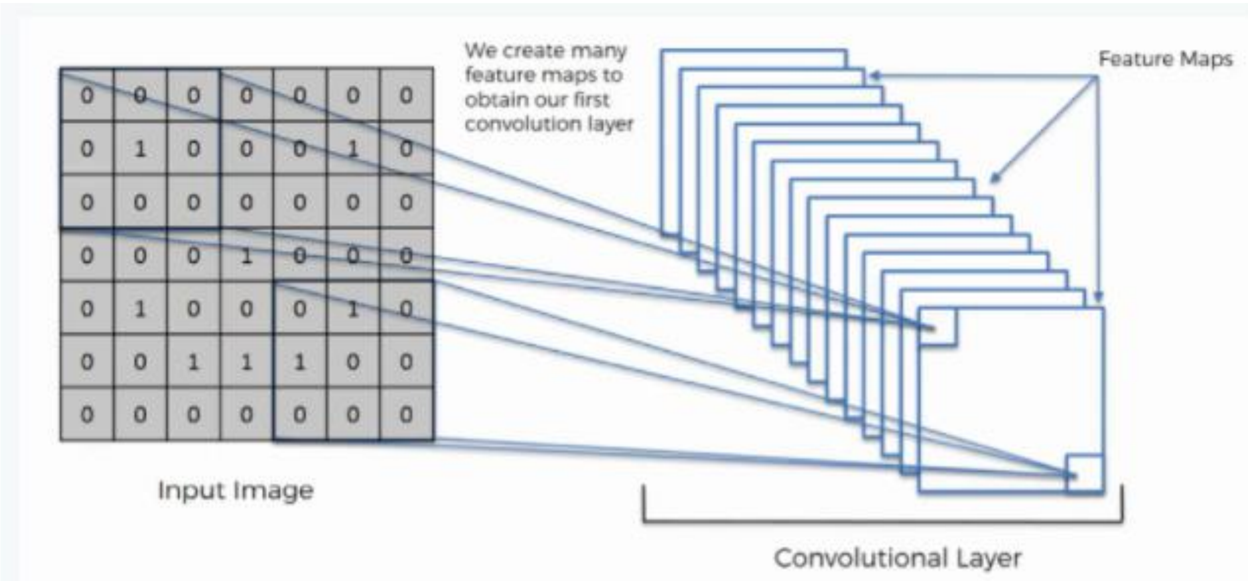


Fig 4.1: Convolutional layer processing of input images

2) <u>Phase Two – Model Optimization</u>

The ResNet34 model is enhanced using the best image size. There are more development options included in order to work on the model's show as well. Twist (0.5) and splendor changes (0.4, 0.7) are among the activities. The next two layers are then prepared and disengaged at the default learning pace. In order to evaluate a progression of learning rates and the number of ages, adjustments are made using this amount and several preliminary tests are conducted.

4.4 Experimental Result and Analysis

The ResNet34 model is enhanced by using the ideal image size. Additional development parameters are introduced in order to improve on the model's performance as well. Twist and splendor changes are among the activities (0.5, 0.4, 0.7). The next step is to prepare the remaining two layers while disengaging them at the default learning rate. This total is used to make adjustments and perform a number of preliminary tests to assess a progression of learning rates and the range of ages.

Table 4.2: Accuracy Table (Train Datasets)

Algorithm Name	**Accuracy (Train Datasets)**
Resnet50	**86%**
CNN	**90%**
Mobile net	**99%**
Vgg19	**99%**
InceptionV3	**99%**
Resnet152V2	**100%**

Table 4.3: Accuracy Table (Test Datasets)

Algorithm Name	**Accuracy (Test Datasets)**
Resnet50	**96%**
CNN	**88%**
Mobile net	**99%**
Vgg19	**99%**
InceptionV3	**100%**
Resnet152V2	**100%**

Resnet152V2 has the greatest accuracy rate, according to tables 4.2 and 4.3. I use the Resnet152V2 computation, and both the train and test datasets are 100% accurate. Because Resnet152V2 offers the most parameters, as I've already described. In this way, the execution will be fast and accurate. At this time, we should get a graphical representation of the dataset preparation and error rates.

RESNET50: Figure 4.2 provides the resnet50 accuracy model.

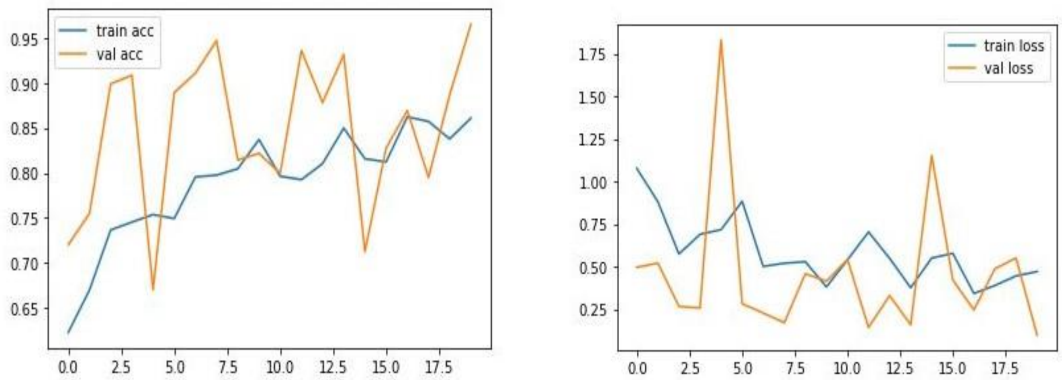


Figure 4.2: RESNET50 Model Accuracy

MOBILE NET: Figure 4.3 provides the mobile net accuracy model.

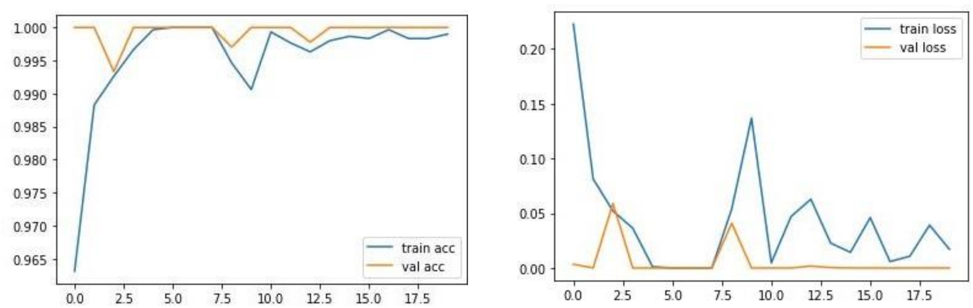


Figure 4.3: MOBILE NET Model Accuracy

VGG19: Figure 4.4 provides the vgg19 accuracy model.

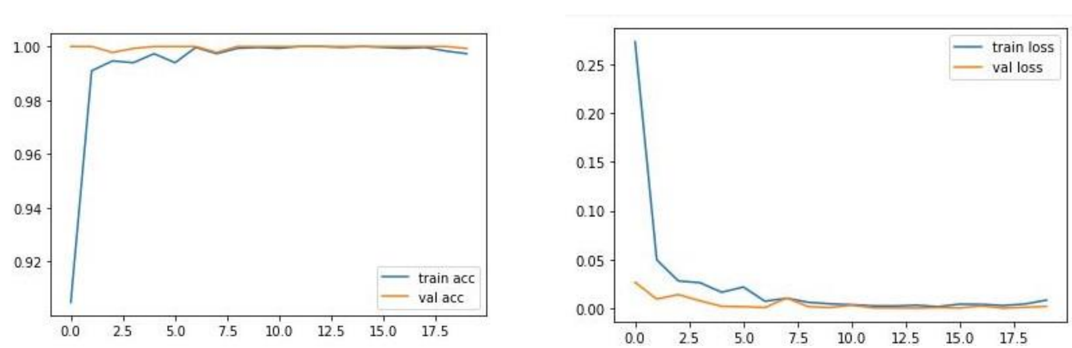


Figure 4.4: VGG19 Model Accuracy

Figure 4.5 provides the accuracy model for inception version 3.

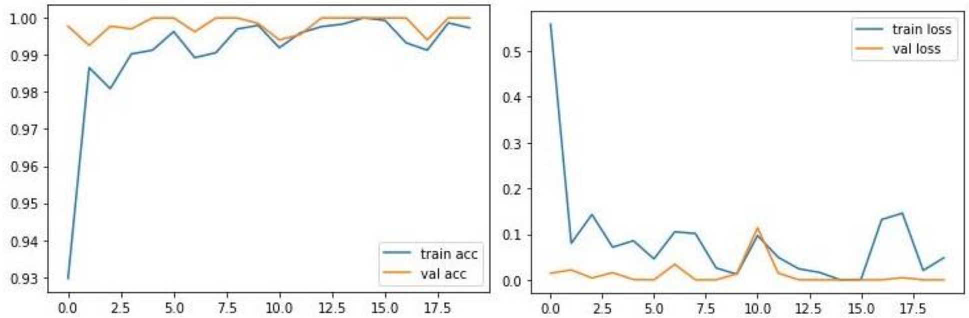


Figure 4.5: INCEPTION V3 Model Accuracy

RESNET152V2: This figure 4.6 shows the resnet152v2 accuracy model.

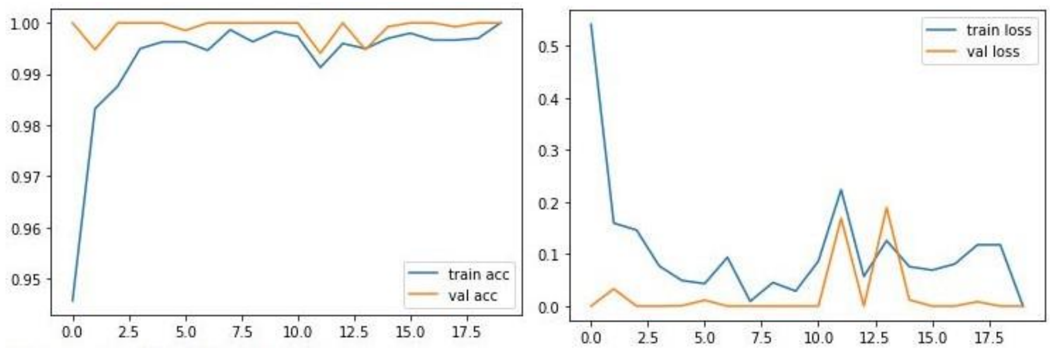


Figure 4.6: RESNET152V2 Model Accuracy

I attempted to utilise test data as the prediction in my model. Using the test dataset, I will identify the water lilies. In figure 4.7, the test code has been merged below.

```
y_pred = model.predict(test_set)

85/85 [==============================] - 10s 117ms/step

y_pred

array([[2.6980083e-11, 1.3255886e-15, 1.0000000e+00],
       [1.0000000e+00, 2.4194638e-13, 3.3591712e-12],
       [8.7483105e-09, 1.0000000e+00, 1.3230704e-11],
       ...,
       [9.9997759e-01, 2.2431332e-05, 2.3611351e-08],
       [2.4272025e-12, 6.3387504e-11, 1.0000000e+00],
       [7.4628888e-12, 1.9041988e-09, 1.0000000e+00]], dtype=float32)
```

Figure 4.7: predicting the test dataset

4.5 Discussion

A model that may be used to identify the origins has been developed by me. However, I get the greatest knowledge regarding the precision level. For my following attempts to be successful, this accuracy rate is essential. Furthermore, accuracy is mostly determined by the dataset. Images are a key component of the study, thus their clarity and quality may be quite high. Furthermore, it really worries me. The algorithms, second. A strong algorithm is essential for accuracy. Since more limits result in greater precision. My precision rate ended up being 90% on the training set and 99% on the test set as a consequence.

The test findings show that although a few best-in-class techniques may have useful applications, other basic approaches, such as innocent representation of the hidden layer output, are inadequate for illness perception. Using highlight representation and semantic word reference, it is possible to extract the visual cues that are often used to describe an illness.

Choosing a representation feasible layer ultimately has a big influence. Even the clarification map, which was made to shockingly detect illnesses, shockingly exhibits features that are different from those in the initial paragraph owing to variances in the organization engineering and the dataset. In light of this, I suggested visualizing each layer and examining which layer is often useful for discernment.

The association of perception techniques revealed the most glaring damage in each picture. The evaluation of the clarity and applicability of the particular methods is engaged by using datasets combined with clarification names for the location of the wounds, which are often created for semantic division activities. CNN could, nevertheless, concentrate on the aspects that I wouldn't anticipate. To prevent over-fitting or dataset tendencies, in such situations, cautious considerations about whether such characteristics have physiological importance should be made.

In overall, CNN's view is that there is a good chance of opening the deep learning mystery. The challenges of using deep learning techniques are less every year, but it's still important for plant researchers to choose the right organization models and understand the outcomes.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

The use of machine learning to predict tomato leaf disease has the potential to have a big influence on society in lots of different ways. First and foremost, producers can prevent widespread crop damage and yield loss by quickly and precisely identifying diseases in tomato plants, increasing agricultural production and ensuring food security. Because it uses less pesticides and has less negative impacts on the environment than traditional chemical-based treatments, this technology also provides an environmentally acceptable option. The availability and affordability of this prediction tool may also provide small-scale farmers in underdeveloped areas more influence, promoting sustainable agricultural and economic development. Additionally, improvements in machine learning for crop disease prediction might open the door for related applications in other fields, technological advances overall and having a good impact on society.

5.2 Impact on the Environment

The use of machine learning to predict tomato leaf disease has a great deal of potential to benefit the environment. This approach can greatly reduce the need for excessive and indiscriminate use of chemical pesticides by enabling rapid and precise identification of diseases in tomato plants. This might reduce the impact of pesticide runoff on the environment and the contaminating of soil and water supplies. Additionally, with the help of the predictive model, focused and effective disease control techniques, there may be a decrease in the total usage of agricultural inputs, further supporting sustainable farming practices and protecting natural resources. Additionally, the strategy can help maintain biodiversity and ecological balance by protecting tomato crops from potential destruction, as well as perhaps easing the strain on land conversion for agriculture. As machine learning techniques develop, their use in disease prediction for crops like tomatoes may serve as a model for comparable environmentally friendly agricultural advancements, aligning with the overarching objective of encouraging sustainable and eco-friendly farming methods.

5.3 Ethical Aspects

It is crucial to consider all ethical implications before using machine learning to forecast tomato leaf disease. As the prediction model would probably depend on large datasets including details about farmers' practices and possibly sensitive agricultural information, data privacy and ownership are an important factor to take into account. Farmers' rights and privacy must be protected by securing this data against unauthorized access and making sure that they have given their informed permission. In addition, it is important to guarantee that everyone has access to the technology and advantages, particularly small-scale farmers in underdeveloped countries who may not have the means to do so. Because farmers must comprehend how choices are made and have faith in the system's suggestions, the machine learning model's transparency and interpretability can give rise to ethical questions. To avoid bias and unexpected consequences, the accuracy and fairness of the model should also be continuously monitored and evaluated. In order to promote trust, equality, and responsible usage in the deployment of machine learning for predicting tomato leaf disease and comparable uses in agriculture, stakeholders must address these ethical issues as this technology develops.

5.4 Sustainability Plan

Several important factors need to be taken into account while creating a sustainability plan for the thesis paper based on a machine learning approach for predicting tomato leaf disease. First off, as new data becomes available and as the agricultural landscape changes, the algorithm should be regularly updated and improved to ensure the predictive model's long-term sustainability. It will be essential to work together with agricultural specialists and stakeholders to spot developing diseases and modify the model as necessary. For the technology to be adopted successfully, encouraging information exchange and capacity building within the farming community will be crucial. Farmers can be empowered to use the prediction tool efficiently and incorporate it into their farming practices by attending workshops, participating in training programs, and using educational resources. Finally, in order to ensure inclusivity and equal advantages, it is important to prioritize the technology's price and accessibility for small-scale farmers.

CHAPTER 6

SUMMARY AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

It focused on utilizing an image from a prepared dataset and a prior informative index to demonstrate how the CNN model was used to predict the example of tomato leaf diseases. The following details on the probability of tomato leaf disease are provided by this. Since the vast majority of leaf types will be covered by this framework, the rancher can learn about a leaf that may never have developed and drills down on all possible plant leaves, assisting the rancher in determining which yields will grow.

5.2 Conclusions

Little-holder farmers rely on an accurate and faultless collect illness diagnosis to avert challenges. In this study, a pre-planned Convolutional Brain Organization was altered, and the model was also posted online. A tomato leaf sickness identifiable proof application was the unavoidable outcome. All you need to utilize this service are a PDA and an internet connection, and it is free. Thus, the needs of the customer as described in this article have been satisfied. The limitations and needs of the model were discovered via a careful analysis. Generally, when authorized in a controlled setting, an accuracy of 94.2% is introduced. This great accuracy is the result of a number of elements, including the disease's stage, kind, foundation data, and object synthesis. For commercial application, a number of client rules would be necessary to guarantee that the specified correctness is delivered. The model was created using a simple base and a single leaf, making it the finest representation of these elements. For the current situation, amplification and move learning were beneficial to the model, aiding CNN in summarizing with more steadiness. This made it simpler for the model to get rid of the highlights, but when the model was given "in field" symbols, it was insufficient.

5.3 Recommendations

Examining vast amounts of data may lead to many different advancements, such as discovering the effects of different activities kids engage in while not studying. The direction of this world is decided by CNN. In Computerized reasoning, CNN is the principal part. Consequently, using CNN to deal with our issues is essential to the progress of the current manifestations for making structure. In CNN, data classifier influences more huge work. In the CNN, the information classifier must be required. The information classifier has the potential to completely alter previous perceptions of the device.

- Utilizing more Pictures information.
- Creating a deeper layer classifier will be beneficial.
- Accuracy increases with more Convolution-layer analysis.
- Technologists and authorities need to pay attention to this industry.

5.4 Implication for Further Study

Farming division needs to robotize the perceiving the yield crops from capability measure (certified time). By demonstrating the assumption, create a web application or workspace application to modernize this cycle to enhance the tasks that must be carried out in the AI system.

REFERENCES

[1] Plant Disease Detection using CNN, Emma Harte School of Computing National College of Ireland Mayor Street, IFSC, Dublin 1, Dublin, Ireland. Published by Computer Tech Science Team, Sunday Sep 28, 2020.

[2] Y.Toda and F. Okura, "How Convolutional Neural Networks Diagnose Plant Disease", Plant Phenomics, vol. 2019, pp. 1-14, 2019. Available: 10.34133/2019/9237136.

[3] Huy Minh Do"Rice Diseases Image Dataset", Kaggle.com, 2020. [Online]. Available: https://www.kaggle.com/minhhuy2810/ricediseases-image-dataset. [Accessed: 26- Apr- 2020].

[4] S.Bharath1, K.Vishal Kumar2, R.Pavithran3, CSE, SRM Institute of Science and Technology, Chennai. India. T.Malathi4, e-ISSN: 2395-0056 Volume: 07 Issue: 05 May 2020 www.irjet.net p-ISSN: 2395-0072.

[5] The Ultimate Guide to Convolutional Neural Networks (CNN), Published by SuperDataScience Team Monday Aug 27, 2018.

[6] Yosuke Toda, and Fumio Okura, Plant Phenomics Volume 2019, Article ID 9237136, 14 pages https://doi.org/10.34133/2019/9237136, Received 22 November 2018; Accepted 11 February 2019; Published 26 March 2019.

[7] Convolutional Neural Networks for Sentence Classification, Published by researchgate August 2014 DOI: 10.3115/v1/D14-1181 Source: arXiv.

[8] S.Bharath1, K.Vishal Kumar2, R.Pavithran3, CSE, SRM Institute of Science and Technology, Chennai. India. T.Malathi4, e-ISSN: 2395-0056 Volume: 07 Issue: 05 May 2020 www.irjet.net p-ISSN: 2395-0072.

[9] Published in: 2018 14th International Conference on Computational Intelligence and Security (CIS) Date of Conference: 16-19 Nov. 2018 Date Added to IEEE Xplore: 06 December 2018 ISBN Information: INSPEC Accession Number: 18343044 DOI: 10.1109/CIS2018.2018.00044 Publisher: IEEE Conference Location: Hangzhou, China.

[10] Chit Su Hlaing; Sai Maung Maung Zaw, 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS) Year: 2018 | Conference Paper | Publisher: IEEE

[11] Conference: Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, April 22-27, 2007, Rochester, New York, USA.

[12] Conference: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009, 6-7 August 2009, Singapore, A meeting of SIGDAT, a Special Interest Group of the ACL.

APPENDIX

RESEARCH REFLECTION 1

I encountered a number of issues during thesis activities. Yet, three issues product major among them. In this paper, I identified some of the major flaws and shortcomings of previous research that used CNNs to identify crop diseases. I similarly gave rules and systems to keep on supporting the ability of CNNs conveyed in obvious applications. Due to a lack of adaptation to some significant AI concepts, some generally distributed arrangements based on CNNs are not currently operational for field use. This shortfall of similitude might incite vulnerable hypothesis capacities for new data tests as well as imaging conditions, which cuts down the valuable usage of the pre-arranged models. Incidentally, the pondered works show the ability of significant learning techniques for crop diseases ID. Their discoveries are unquestionably encouraging for the development of new farming tools that could contribute to a food production that is more manageable and secure.

Machine Learning Approach to Predict Tomato Leaf Disease

ORIGINALITY REPORT

16%	15%	7%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	8%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	7%
3	Nazmun Nessa Moon, Ms. Shayla Sharmin, Refath Ara Hossain, Israt Jahan, Fernaz Narin Nur, Md. Mosfikur Rahman. "Deep Learning Model for Detecting and Diagnosing Plant Disease", 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), 2021 Publication	1%
4	ironset24.com Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		