TOMATO LEAF DISEASES CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK

BY

KAZI ABU SAYEM MD. NAYEM
ID: 191-15-12925

KAJALI CHAKMA
ID: 191-15-12923

A.S.M MONIRUL HASAN
ID: 191-15-12919

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

Supervised By

LAMIA RUKHSARA
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

DR. MD. TAREK HABIB
Associate Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Research titled **"Tomato leaf Diseased Classification Using Convolutional Neural Network"**, submitted by **Kazi Abu Sayem Md. Nayem, ID: 191-15-12925, Kajali Chakma, ID: 191-15-12923, A.S.M Monirul Hasan, ID: 191-15-12919** Student ID to the Department of Computer Science and Engineering, Faculty of Science and Information Technology, **Daffodil International University**, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc in Computer Science & Engineering and approved as to its style and contents. The presentation has been held on 23th January 2023.

BOARD OF EXAMINERS

Chairman

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

Internal Examiner

Dr. Md. Zahid Hasan
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Fahad Faisal
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

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 **Lamia Rukhsara, 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:

Lamia Rukhsara 23.01.23

Lamia Rukhsara
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Dr. Md. Tarek Habib
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Kazi Abu Sayem Md. Nayem
ID: 191-15-12925
Department of CSE
Daffodil International University

Kajali Chakma

Kajali Chakma
ID: 191-15-12923
Department of CSE
Daffodil International University

Monirul Hasan

A.S.M Monirul Hasan
ID: 191-15-12919
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we would want to give God the highest praise and gratitude for His wonderful grace, which enabled us to successfully finish the senior thesis.

We would like to express our sincere gratitude and debt of gratitude to **Lamia Rukhsara**, a senior lecturer in the department of computer science at Daffodil International University in Dhaka. We were able to successfully finish our assignment on "Image Processing" thanks to our in-depth knowledge, intense curiosity, and helpful directions. This endeavor was made possible by her never-ending patience, academic guidance, constant encouragement, frequent and vigorous supervision, constructive criticism, invaluable advice, reading many subpar drafts and fixing them at all stages.

We would like to extend our sincere gratitude to **Dr. Touhid Bhuiyan**, Head, Department of Computer Science and Engineering, Faculty of Science and Information Technology, DIU, for his invaluable assistance and counsel in helping us complete our project. We also extend our sincere gratitude to the other professors and employees of the department of CSE at Daffodil International University.

Finally, we would like to express our gratitude to all of the well-wishers, friends, family, and elders for their support and inspiration. Hard labor, as well as all of those inspirations and help, went into this study.

Finally, we must respectfully appreciate our parents' unwavering assistance and endurance.

ABSTRACT

Rapid and precise disease detection can considerably boost the sustainable agricultural yield of tomato plants. The conventional method allows for human agricultural professionals to recognize irregularities in tomato plants caused by pests, sickness, bad weather, and nutrient shortages. Initially, less precise traditional processing of image and artificial intelligence approaches are being used to automatically identify diseases in tomato leaves. To improve accuracy of prediction, deep neural network-based categorization is implemented. This article introduces a thorough analysis of current research on the tomato leaf detection diseases via image processing, machine learning, and deep neural network methods. Discuss the methods employed and deep learning frameworks that have been adopted, as well as the public and private datasets that may be utilized to detect tomato leaf disease. As a consequence, suggestions are made on how to choose the best approaches in order to increase forecast accuracy. The complexities faced when a the machine learning and deep neural network models are then outlined.

TABLE OF CONTENT PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Among the most important human endeavors that promote national growth is agriculture. Due to population increase and the need to meet their food needs in order to improve their quality of life, there have been significant developments in the farming and food industries recently. Most of the nation's economic activity is dependent on agriculture, which is not only the ingredients and food. Plant diseases can endanger the safety of the world's food supply and be a catastrophe for small farmers who lack the tools to combat them. One of the most important aspects of managing crop disease overall is taking preventative action by spotting plant illnesses in their early stages in order to prevent further damage. Methods for detecting plant diseases based on automated object identification may be implemented using the technology of cell phones, which really is being assisted by fast increasing processing power. In a number of research, plant disease identification has been aided by machine learning. There are multiple Machine Learning (ML) models that have been used for picture detection and categorization. However, classic picture recognition techniques have a weak capacity to generalize. This field of DL study seems to have a lot of potential in terms of improved accuracy given the ongoing advancement of DL technology. Researchers used DL architectures for image recognition and classification as they began to advance over time. Additionally, these structures have been used in a variety of agricultural applications.

Traditional machine learning approaches often require hand-crafted features as inputs to the algorithms. It could be challenging to apply this approach to online applications since it calls for more sophisticated computations. Additionally, the performance obtained might not be adequate. It may be necessary to extract characteristics using more advanced tools, including spectroscopy [3], spectrum patterns in the infrared and crop phenotyping [4], in order to achieve higher accuracy. The tools needed to extract the traits, meanwhile, could be out of reach for smallholder farmers.

1.2 Motivation

Our major goal is to use smartphone applications to identify ailments to classify tomato diseases. Additionally, we can quickly and easily identify illnesses in tomatoes, leaves, and other plant materials thanks to machine learning techniques.

- Demonstrating the feasibility of using deep convolutional neural networks to classify plant diseases.
- Developing a model that can be used by developers to create smartphones application to detect plant diseases.

This study definitely aims to specifically locate utilizing crop diseases robotic platforms. Finding pests and illnesses really is necessary to prevent harm to plant kind of leaves. These harmful impacts affect the leaf"s generally physical appearance, making it easy to kind of pinpoint the damage\'s origin from photos taken by cameras. The use of a mobile computer and an everyday RGB camera really is necessary to use a robot to kind of detect illnesses in a subtle way. Robots for the most part are used to particularly identify illnesses, thus a mobile computer must be able to for the most part do computational duties. Other tasks that mobile computers may basically do include image processing, driving, and autonomous navigation, which is fairly significant. Detection algorithms must thus mostly be basically quick and efficient in terms of essentially RAM utilization, which is fairly significant.

1.3 Rational of the Study

The process of segmenting consists with breaking apart the image of leaf texture into more manageable pieces. Using the prescribed method, the tomato leaf's border is drawn, and the leaf is then cut into sections and labeled. The segmentation process is stopped by using the similarity and dissimilarity of two pixel intensity traits. Find similarities using color-based thresholding. To achieve this, choose the foreground pixels from the tomato picture's natural intensity range.

1.4 Research Questions

Our capacity to do this assignment was put to the test. The following questions have been put forth by the researchers in an effort to realistically, correctly, and effectively convey the thoughts and impacts of the event.

- Is the agricultural industry still one of the most crucial industries on which the majority of the people of Bangladesh depends?
- Is the architecture used to divide tomato leaf diseases into 10 categories a straightforward convolutional neural network with the fewest possible layers?
- Is it necessary to lessen the human labor required for this activity, while still producing reliable forecasts and ensuring that the farmers enjoy stress-free lifestyles?

1.5 Expected Output

Given that focuses were the primary outcome we were hoping to see, there are a few focuses in this section. The expected result of this investigation-based research study is the creation of an method or comprehensive, effective process that will categorize tomato leaves in accordance with the trained dataset's built-in model. It is challenging to detect and extract crucial data from tomato leaves that might assist in differentiating between the traits of various ailments using traditional image processing techniques. Due to the wide variation in these disorders' characteristics, a detailed automated investigation of their patterns is required. The only variables influencing how well the machine learning-based models perform are those of the manually selected, created features. In order to choose and learn the ideal set of characteristics for classification, feature extraction must be automated.

1.6 Project Management and Finance

Our project is basically a research base project. Here we have collected data from Ministry of Agriculture and Kagal. In this case we have no financial issues. For collect the data from agriculture ministry, some money was spend and nothing else has been spend so far. But, in future we may have to spend some money to develop our project. For better collab function we may have to buy a subscription to upgrade our research project features. Despite the fact that the program is produced successfully and the cost was quite little and in expensive.

1.7 Report Layout

With the venture's aim, motivation, research questions, and projected outcome established in Chapter One, this part outlines the report's overall structure.

Chapter 2 discusses previous work that has been done in this field. The scope that resulted from their field's constraint is then demonstrated in this second chapter's subsequent portion. Finally, and most importantly, the main obstacles or difficulties in this research are described.

The theoretical debate around this research project is described in Chapter 3. This chapter elaborates In this study, statistical methods were employed to address the theoretical component of the study. Additionally, this chapter shows how CNN and a artificial intelligence classifier use procedural techniques. There is a supplied confusion matrix analysis in the chapter's last section to approve the presentation and make the precision name of the classification visible.

The study findings, a presentation analysis, and a conversation of the results are presented in Chapter 4. This section includes a few experimental images to help the project come to life.

The topic of chapter five was summarized along with future work and a conclusion. The whole project report is expected to appear in this chapter after recommendations. The chapter is concluded by pointing out the limits of our work, which may affect other people who need to work in this subject in the long term.

CHAPTER 2

BACKGROUND

2.1 Preliminaries

The research challenges, pertinent papers, and a research overview are presented in this section. The section on comparable works will discuss other research publications as well as the methods, accuracy, and works that are pertinent to our work. The section below will include a summary of the study. Our companion pieces in the section on problems, we'll look through how we improved accuracy.

2.2 Related Works

To successfully move in the right path, it is crucial to acknowledge the prior study that has been done in this sector. To accurately conduct the classification crop leaf diseases, image processing and deep neural network techniques have been widely applied in following important research ground. In this essay, we analyze the methodologies that have been used the most frequently in the literature in the topic as well as wholesome tomato leaves. If done manually, keeping track of a sizable field of crops is a laborious operation. It is essential to reduce the amount of human labor required for plant management. As a result, this is a prominent study area that draws several researchers. In literature, plant diseases are mentioned in a number of works.

An effective approach to determine if a tomato leaf is healthy or sick has been put forth by the authors of the study [7]. The image that was provided as input was being processed beforehand by erasing the background and with use of any noise loss method. The improved image's texture features extraction with the use of Gray Level Co-occurrence Matrix (GLCM). After being trained with a variety of kernel functions, the Support Vector Machine (SVM) classifier's performance was evaluated using the N-fold cross-validation method. The proposed system's accuracy rate while using the Svm classification and linear model is 79.83%. Although excellent, the accuracy gained is unsatisfactory to predict or differentiate between good and ill leaves.

The authors of [13] have presented several feature withdrawal, separation and classification algorithms that recognize and notice the kind of the illness utilizing the sick picture to do sorting in order to address the issue with the aforementioned research. The leaf picture which was provided as the structure's input was pre-managed, either by making it smooth or by improving it using histogram balance. Diverse separation methods, such as KM clustering has been used to pinpoint the area that is most affected. The segmented region's structures were then removed, and GLCM calculations were made utilizing them. After feature extraction, Synthetic neural networks may be used to identify the disorders. The technique was only partially automated because the user had to specifically choose the cluster that included the sick portion when segmenting the picture using K-Means clustering.

MATLAB has been used to implement certain research projects [11] [12]. To attain the necessary detection levels. It is necessary to develop and implement statistical measures for this detection system that are processed in terms of reactivity, specificity, and accuracy in a much more challenging, consumer setting.

For photos captured by field instruments or remote sensing equipment, image processing is a common practice in precision agriculture. Image processing in agriculture was a common topic in literature. For instance, image processing is utilized for weed identification, fruit grading, and the detection, quantification, and classification of plant illnesses [51, 61, and 71], as well as the phenotyping of signs of plant diseases. Deep learning has been applied to detection recently. In order to identify illnesses from the leaves of diverse plants, Mohanty et al. used deep learning in their research [81].

A technique to identify leaf disease using hyperspectral measurement was described by Ashourloo et al. [101]. In [111], a method for assessing the severity of the illness in leaves was put out. For calculating the severity level, statistical characteristics in the RGB and HSV color spaces were applied. Otsu's segmented and decision trees for categorization are combined, H. Sabrol et al. demonstrated a method for tomato leaf illness finding. For studying the features of the leaf illnesses, our method took into account color, shape, and texture parameters [121].

In order to identify 26 diseases in 14 distinct plant species, convolutional deep neural networks have been established in [65]. For it, the authors adopt widely used GoogleNet [105] and AlexNet [45] designs. This has been accomplished by using a public resource that comprises 54,305 pictures of both damaged and healthy plant leaves. Images of crop leaves was being collected in a precise setting to shape the data.

The authors of [15] provide a method for identifying and categorizing the banana sigatoka and banana speckle illnesses that affect banana leaves. They worked under some difficult circumstances to develop deep learning models. These circumstances include lighting, a complicated background, various image resolutions, sizes, and orientations. They successfully illustrate the precision of this method and the negligibly high computing effort needed.

2.3 Comparative Analysis and Summary

Machine learning may be completed via profound learning. It is constructed using synthetic neural networks. Neural networks function similarly to human brains. Convolutional neural networks, or CNNs, are among the most well-established methods for considerable learning. It's a phony neural mastermind, sometimes referred to as a feed-forward ANN. The frameworks receive information immediately in a "feed-forward" sort out.

CNN was created by Yann LeCun. He created it, inspired by human architecture. CNN really does function like the normal visual cortex. One of the most successful photo grouping models is CNN. The grouping precision of CNN is superior to certain other standard picture order estimates. In CNN, include assurance is not required, but it is required for other picture characterization calculations. On CNN, many kinds of layers are applied. An image-passing moving channel or bit is incorporated into the convolution layer. It moves through a 2D arrangement (a representation of the image) all around, takes a particular package, applies bit duplication, and saves it in a different lattice. Researchers like Chen Introduced better versions of CNN like EfficientNet to make the pursuance easier.

2.4 Scope of the Problem

We have studied a great deal of research papers and have learned more about our methods for solving problems. However, we utilize the best model, effectivenetb3, more frequently than other models. When we first began implementing it, we encountered a few challenges, but after consulting earlier research papers, we were able to resolve them. Finally, we succeeded.

2.5 Challenges

There are still some significant challenges to overcome in the finding of illness in tomato leaves despite multiple approaches.

- It is somewhat challenging to recognition and extraction of crucial information from tomato trees to differentiate the traits of distinct illnesses using old-style picture dealing out techniques. It is vital to conduct in-depth study on the patterns of these diseases utilizing a range of automated datasets due to the high heterogeneity in their characteristics.
- The sole variable that affects how well the artificial intelligence-based models perform is the type of the personally chosen and made characteristics. In order to choose and learn the ideal set of characteristics for classification, feature extraction must be automated.

- Most deep learning models provide a same weight to each feature produced from each layer. To improve the model's sensitivity for sorting, nonetheless, every tier must do features weighting. As a result, significant traits may be found and sent onto network's deeper tiers for more accurate categorization.

- The deep neural network must also are trained using a lot of examples to achieve improved feature simplification.

.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

We're striving to make our ground-breaking system for classifying tomato diseases in this area more transparent. The gathering and evaluation of data, the description of the proposed model with its major assumptions, the chart, table, and explanation are some of the key components. claim to have created the Connected television show for CNN and to be the owner of the dataset utilized in this study. An description of the numeric concepts is given in the chapter's conclusion that underlie our project in addition to offering a complete comprehension of the finishing necessities.You may consider the subject of inquiry to be a study area that was looked at and researched to elucidate ideas. For execution and model building, data gathering, implementation, or processing, and model training. We talk about the technology and methods we used under instrumentation. The Python libraries numpy, pandas, skit learn, matplotlib, seaborn, tensorflow, shutil, etc. were all utilized together with the Windows platform. Colab by google, a free and Python distribution for use in information science and artificial intelligence requests, was used for all of the implementing and validating.

3.2 Data Collection Procedure

The technique used in this investigation involves several steps, including data collection, data process, data resize and augment, method selection, etc.

Stage 1 - Data Collection: In order to create our own data collection, we processed the raw data that we obtained from the Ministry of Agriculture and Kaggle. There isn't a single dataset in this subject because gathering data was so difficult.

Stage 2 - Data Processing: Each class produced its own information after compiling data from numerous sources. There are a lot of unreliable and noisy data. Before implementing the chosen dataset, we first manually process the relevant data.

Stage 3 - Data Resize and Augmentation: Following processing, data have been lengthened and reshaped, class by class. For training reasons, we had to go through data augmentation and scaling. We only applied the most important augmentations due to the potential overfitting of enhanced data.

Stage 4 - Model Selection: In order to improve the accuracy of our data, we choose a model to train and test. Many people utilize convolutional neural networks. One model was finally picked for the training and testing stage after we applied many models to enhance the accuracy of our machine configuration.

Stage 5 - Performance Evaluation: In this part, all of the results are presented as graphs. These procedures gave us a few accuracy graphs indicating accuracy and validity reduction during training and testing. Additionally, we produced a table that shows the confusion matrix, f1 score, accuracy, and recall metrics.

Stage 6 - Conclusion and Future Work: This section will provide a summary as well as a schedule for more development.

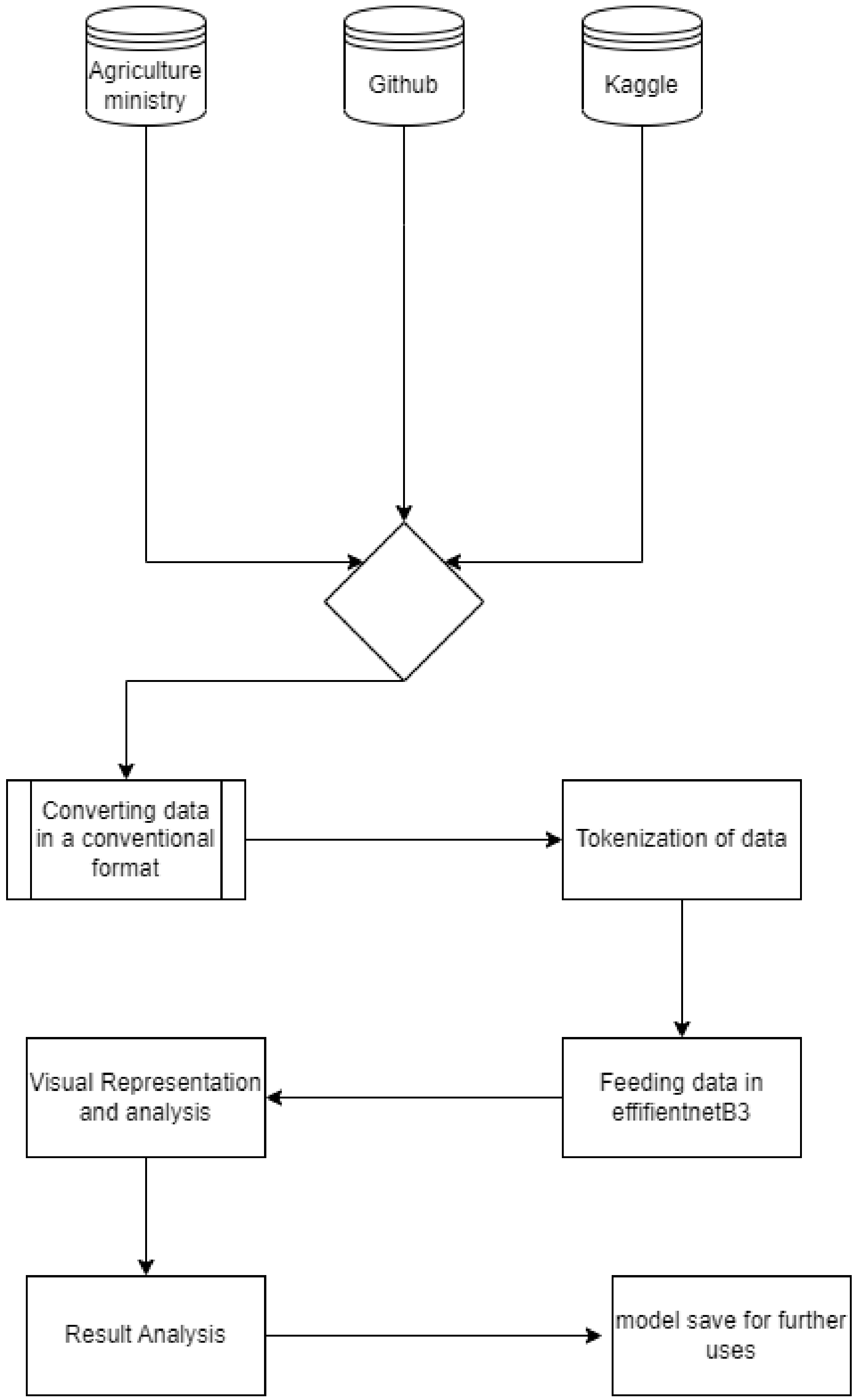

Fig 3.3.1: Workflow of our approach

We processed the unprocessed data that we acquired from Kaggle and the Ministry of Agriculture in order to produce our own data collection.

Each of the four different tomato leaf varieties has 2000 different photos in a Bangladeshi database that we created. Our primary data source, which contained 9000 photos, was the Kaggle dataset.

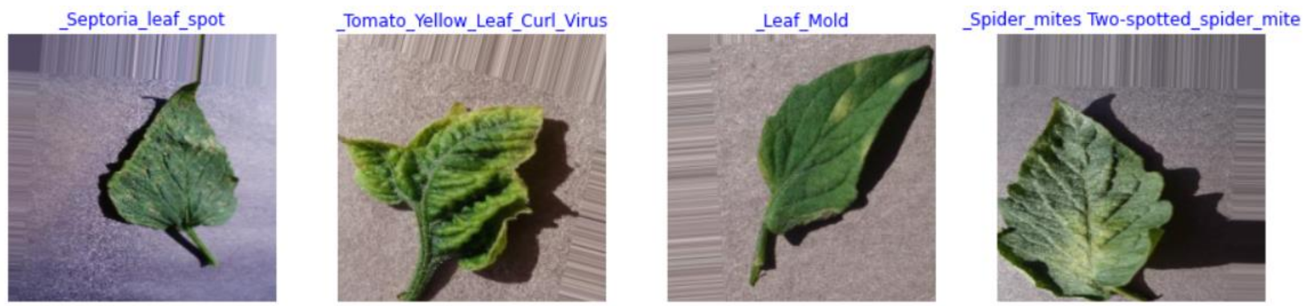

Fig 3.4.1: Images of tomato leaves from the Kaggle Database.

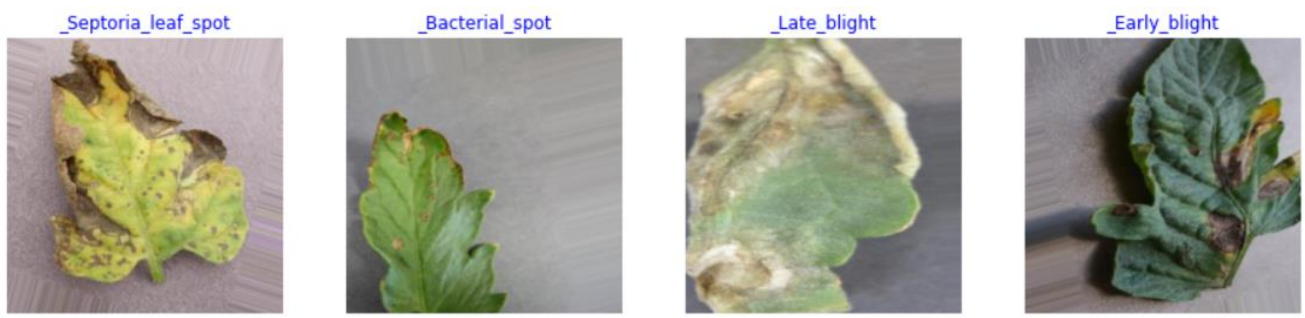

Fig 3.4.2: Images of tomato leaves from Bangladeshi database.

3.3 Statistical Analysis

A two-step data handling mechanism Information planning and information augmentation are two examples. The outcome of our negotiations with push information often hinges on the preprocessed data. Pre-processing information more efficiently will lead to more exact results. In a single sentence, it is the initial hurdle for this type of inquiry-based activity.

3.3.1 Data Augmentation

To keep a strategic distance from overfitting, we deceptively expanded the dataset. By incorporating data from both internal and external sources into a project, it adds value to the information on which it is based. Extending the quantity of meaningful information inside the dataset makes a difference.

We provided the greatest detail in five different approaches., these approaches given here:

- Rotate left -30 degrees: Rotate left -30 degrees enhancement is used to enhance the data set.
- Mirror:The most often used augmentation to increase the dataset is the mirror augmentation.
- Shear: We utilize this method to do shear amplification.

$$A=\begin{pmatrix}1 & s\\ 0 & 1\end{pmatrix}$$

3.3.2 Data Preparation

All of the images in our collection are different heights and widths. We scale our dataset into 256 x 256 pixels because our demonstration demands a consistent pixel for every image.

Additionally, we converted the images to grayscale. Since our computer's graphics processing unit may have been better, we used that computer to prepare the show. So, in order to create the model, we used grayscale pictures.

3.4 Proposed Methodology

We offered a method for handling tomato leaves. We first developed single-stage tomato leaf localization and two-stage tomato leaf construction using a monocular image. We developed an architecture for a neural network that can recognize various leaf diseases. In our investigation, a dataset comprising 10 distinct forms of leaf disease was used.

3.4.1 Neural Network Architecture

We proposed an architecture that uses EfficientNetB3 with noisy initial weights and four double convolution blocks operating concurrently. Utilizing two stages of upscaling and a convolution layer as the head, we merge features.

We advise trimming the RGB input image to remove the higher area that contains unnecessary elements. The image is then submitted to EfficientNetB3 to perform four double convolution blocks of parallel feature extraction. There are two outputs from the double convolution blocks:

(1) An upscaling layer is used to blend the first one with EfficientNetB3.
(2) The another upscaling layer adds the production of the first to a shallower block of the second convolution. To provide a balanced output between nearby and distant leaves, upscaling is used.

3.4.2 EfficientNetB3

EfficientNet, which was first introduced in Tan and Le in 2019 and uses the fewest FLOPS for inference, is one of the maximum efficient algorithms and accomplishes State-of-the-Art accuracy on mutually commonly used image categorization transfer learning problems with imagenet. EfficientNet is ImageNet's most recent state-of-the-art technology. It offers an almost perfect way for scaling CNNs (Convolutional Neural Networks). EfficientNet, a well-known design, was able to solve a number of machine learning problems in the most innovative way. It uses EfficientNet to perform a range of photo categorization tasks. It is a suitable model for transfer learning because of this. Using that as an illustration, we'll illustrate using EfficientNetB3, which has already been trained.

3.4.3 Convolutional Layer

CNNs are widely used in photo and Natural language processing, recommender systems, and video recognition. networks of neurons with convolutions. These systems may sound like an odd blend of physics, math, and computer science with a dash of CS, but they have been some of the most influential developments when it comes to machine vision. The year when artificial neural network gained notoriety was in 2012, when Alex Krizhevsky used them to win the ImageNet competition (basically, the annual Olympics of computer vision), lowering the classification error record from 26% to 15%. At the time, this was a startling achievement. A cluster of pixel values is what a computer sees when it receives an image as input. The image may determine and estimate that a 32 x 32 x 3 cluster of numbers is there (The 3 alludes to RGB values). There are more layers layered in between these convolutional layers in a traditional convolutional neural network architecture. I would strongly encourage everyone who was interested.

3.4.4 Feature extraction

CNN's greatest important structure blocks is convolution. Convolution is a phrase used to describe the scientific fusion of two abilities to harvest a 3rd work. It combines two pieces of data. When we talk about CNN, the density is carried out on the contribution data using a channel or portion to eventually produce a highlight outline. Convolution is carried out by moving the station over the input. A system repetition is carried out at each location, and the result is added onto the highlight outline.

3.4.5 Max pooling

Convolutional Neural Systems' down testing method might be Max Pooling. The objective is to lessen complexity of an input representation by downsampling in order to make assumptions about the highlights in the binned sub-regions. As a result of the larger amount of pixels contributing to additional constraints, which might include significant amounts of info, it is mostly used to lessen the scope of the picture. This reduces the amount of parameters needed so that CNN can still differentiate the image. 75% of the enactments and overfitting are eliminated by max pooling.

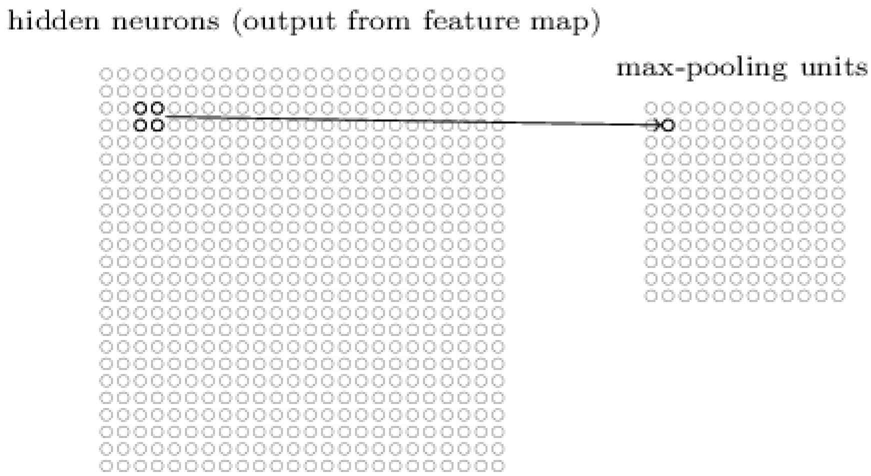

Fig 3.6.3: The highest activation output comes from a pooling unit..

3.4.6 Dense Layer

Another name for the fully linked layer is a dense layer. Similar procedures are carried out at the thick layer where every neuron is connected to every other neuron. Since it refers to a thick connection of thick neurons, it is also known as dense. Weights associated with each neuron combined with unique values are included in a thick layer. For high-level reasoning within the neural network, several types of work, including SVM, softmax enactment work, and many more, are used here. However, we adhered to using softmax for categorization in our demonstration. We receive a few high-level highlights as input after performing a few convolutions and pooling layers. These input picture highlights are used in the categorizing process to look into several classifications. However, when we mix the highlights from the surveying layer and the convolution layer, the classification results are much improved. Sum of yield probabilities in completely associated layers is 1. Weights from one Conv layer are shared with other Conv layers. Using a softmax layer to connect all hubs is really challenging.

3.4.7 Soft-max

Let's think about a categorization model that uses n classes. This model generates a score for each class using input datasets and an algorithm. The likelihood between 0 and 1 is converted into score via the softmax activation function. All probability added together have a value of 1. To categorize the classes, we used this research to the top coating of CNN. Several lessons are offered by this work from a variety of inputs. the likelihood dissemination of softmax work is:

$$\sigma(x_j) = \frac{e_{xj}}{\sum_{j=1}^{n} e_{xi}}$$

Where i =1, 2, 3 ……. n and j =1, 2, 3 …….n

3.4.8 Dropout layer

Dropout might be a technique used to advance over-fit on neural networks. A neural network's data structure might be a coordinated chart with each hub representing an inclination and each edge representing a weight. When the signals that enter y through edges e1, e 2, and e 3 are each received independently and there is unity with predilection b and weights w1, w2, and w3, the yield of y is w1x1+w2x2+w3x3+b. We are aware that an arrangement has more parameters the more complex it is. As an example, the 2014 ImageNet competition's VGGNet includes a mere 148 million parameters. That's distinct. With that many factors, the arrangement may easily overfit, especially if the dataset is small. Since CNN must operate in a demanding climate, dropout becomes essential. Dropout is mostly selected between 0.2 and 0.8. Dropout randomly empties the nerve cell based on the client-provided constraints, such as 0.4, etc.

3.4.9 Flatten Layer

Utilizing this will be necessary in order to later incorporate the data into a fake neural network. Unlike convolutional layers, completely associated layers do not have a proximity limitation (which, using convolutional channels, was observing a few close areas of an image). This suggests that it can merge every neighborhood highlight discovered by the prior convolutional layers. A "straightened" 2D cluster might be created for each highlight outline channel in a CNN layer's output by combining the outcomes of several 2D parts, one for apiece channel in the contribution layer.

3.5 Implementation Requirements

The Kaggle Database analyzes 9000 images of tomato leaves that were captured from various points of view. 2000 unique photographs may be found in the Bangladeshi database that we created. 8000 photos from acquired data were used to train the model, and 3000 pictures were used to validate the method. We utilized 20% of the dataset to test and validate the dataset after using 80% of the dataset to train the model. The network has been set up for five epochs.

After carefully examining all pertinent statistical or theoretical concepts and techniques, a list of requirements for such a classification assignment has been developed.

The following are likely essential items:

Hardware/Software Requirements

Operating System (Windows 7 or above)

Hard Disk (minimum 500 GB)

Ram (Minimum 4 GB)

Developing Tools

Python Environment

Google Colab

Jupyter Notebook

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

We illustrated the growth of many tomato leaf varieties in this location. They often break down the preparation of the demonstration into a small number of processes, such as dataset gathering, material arrangement, information expansion, material resizing, and proposal.

4.2 Experimental Results & Analysis

When the demonstration is related to the preparing information, preparing precision is typically the exactness. Approval accuracy occurs when the demonstration is linked to randomly chosen images from the individual course. A chart containing the preparation and approval exactness of our technique can be shown in Fig.

Fig 4.2.1: Training and validation accuracy

The error of gathering a collection of knowledge results in training loss. The error that results from putting the approval set of data via the prepared network is known as approval misfortune. A chart containing the planning and approval loss of our demonstration may be seen in Fig.

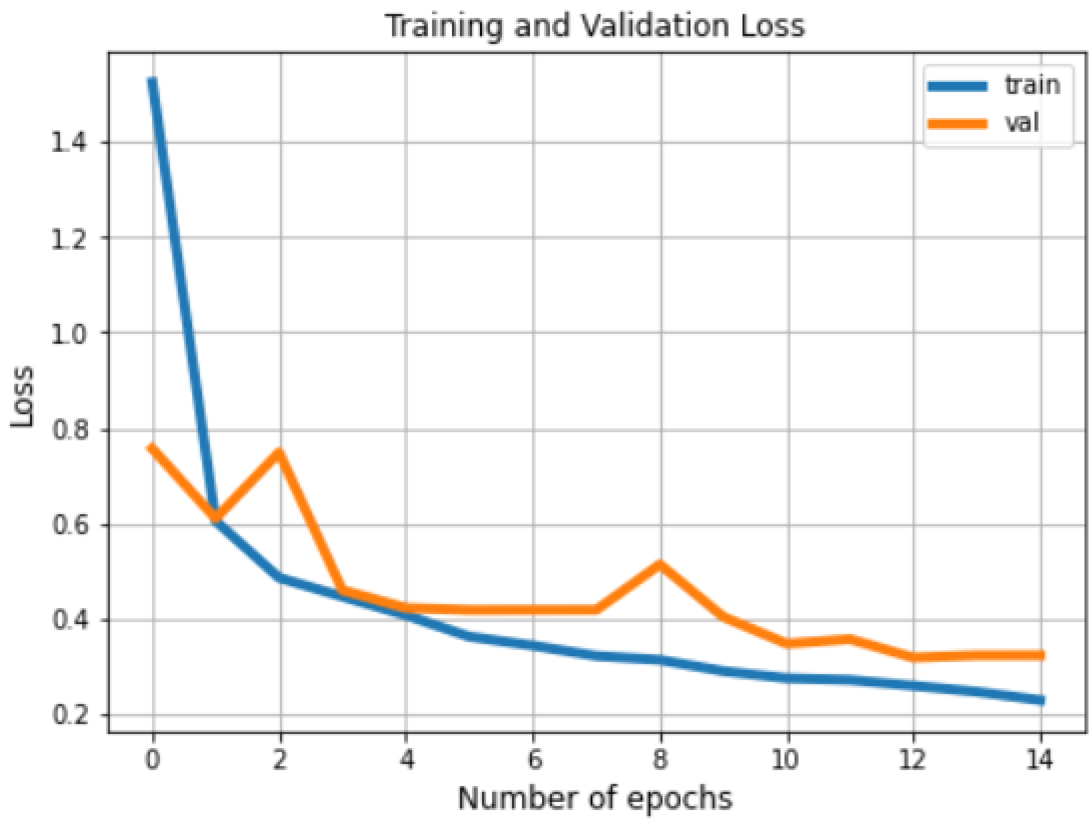

Fig 4.2.2: Training and validation loss

Comparison table:

Serial no.	Model	Accuracy
1	Mobilenet	63.75%
2	VGG 1	77.2%
3	Inception V3	63.4%
4	Proposed model(EfficientNetB3)	96.10%

Table 4.2.2: comparison of different model

4.3 Discussion

In this experiment, we used a test dataset of 11,000 images to measure precision and recall. Precession normal is 0.00 and recall normal is 0.00, according to the order report. So, it is really possible to say that our classifier works fantastically and outstanding. It is possible to observe that the classifier had a respectable level of precision from the table of the classification report that was created.

Precision: The section of retrieved reports that are focused on precision which is relevant to the inquiry within the data recovery industry:

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

The percentage of all relevant archives that the lookup returns is used with precision. Keep in mind that the notion of "precision" and how it is used in the field of data recovery are different from how exactness is defined and used in other disciplines of research and innovation.

Recall: Recall is the portion of important events that have been identified among all important occurrences. The tall recall suggests that the majority of the relevant result was returned by a computation.

$$Recall = \frac{tp}{tp + fn}$$

Accuracy: The similarity of the calculated values to a known standard is referred to as accuracy.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on society : Plants are essential to our existence because they give us food and shield us from radiation. The tomato is a food-rich plant that is frequently grown as an eating vegetable. Life cannot be conceived without plants.

In this instance, tomato leaf disease is a crop-threatening issue for tomatoes. Because of this, the tomato plants are unable to flourish, and the output is likewise poor. Even farmers have financial difficulties. These diseases are cursed if they are misdiagnosed. Illness control strategies could be ineffective and worsen plant damage if the disease and its source are not correctly identified.

5.2 Impact on Environment

By utilizing technology, farmers may increase tomato productivity while preventing tomato leaf disease and producing high-quality tomatoes. By getting rid of those tomato pests, tomato leaf blight can be avoided. Tomato leafhoppers will therefore vanish from the environment.

5.3 Ethical Aspects

According to the concept of informed consent, participants in a study must have complete access to all information pertaining to the study before deciding whether or not to take part in it. This standard states that the study must have enough scientific value to convince volunteers to incur study risks. Confidentiality: In accordance with this rule, researchers are required to make all reasonable steps to protect data and make sure that private information is kept private.

According to this concept, the study's benefits must balance its drawbacks. The risks of not detecting a plant disease sooner include the possibility of global hunger, impairment to the agricultural industry, and harm to everyone's health in human civilization. All of the aforementioned principles were successfully applied in our study.

5.4 Sustainability Plan

According to this requirement, the study must have sufficient scientific value to persuade volunteers to accept study risks. Researchers must take all necessary precautions to safeguard data and ensure that sensitive information is kept private in compliance with this regulation. Therefore, proper diagnosis is necessary.Our research will help the farming community learn about these issues and increase their enthusiasm and confidence in machine learning.

CHAPTER 6

CONCLUSION, RECOMMENDATION AND FUTURE WORKS

6.1 Summary of the Study

A trustworthy approach for determining the kind of contagion in tomato leaves is provided by this study. This, in our opinion, is the first attempt to identify tomato sickness using the attention gating machinery in enduring CNN. The attention mechanism that is added to the Lasting network for effective feature learning is the main contribution of this study. It is beneficial to selectively balance the characteristics of several layers at the beginning of a single layer. As a consequence, the receptive field at a layer is increased in order to view feature maps from different levels of the processing architecture. The current layer can now include more contextual data into the input. Currently, learning at the layers underneath it is aided by perception of the properties at the current layer. This is caused by the back broadcast of the tensors sideways the omission networks.

6.2 Conclusion

The projected method has a 95% accuracy rate. The recommended method uses a CNN process for graded feature withdrawal to map contribution picture pixel strengths and comparation them with the training dataset image. All modifiable leaf component parameters are improved by lessening error in relation to the train dataset. To differentiate between leaves damaged by illness and healthy leaves, an image classifier technique is used to identify the comparative image. This approach may also employ artificial neural networks, fuzzy logic, and hybrid algorithms. The recommended work may be further developed to incorporate a number of cutting-edge algorithms in order to provide the greatest results in comparison to present approaches. The categorization of plant diseases based on real-time applications would be one of the important factors in selecting the best strategy.

6.3 Implication for Further Study

The farming area, being one of the most vital, continues to be strongly dependent on the majority of Bangladeshis. Consequently, identifying diseases in these crops is crucial for the economy to grow. The tomato is one of the main crops that is widely grown. The aim of this article is to identify and describe 10 dissimilar illnesses that affect the tomato harvest. The recommended method uses a CNN model to order tomato leaf illnesses discovered in the Kaggle dataset. A simple convolutional neural network with a small amount of layers was used as the architecture to divide tomato leaf sicknesses into 10 distinct modules. Future research may also entail testing the proposed model with numerous erudition rate and optimizer. It may also entail tests with more contemporary constructions to progress the performance of the method on the train set. In order to aid and backing them in finding sicknesses that may impact tomato plants, farmers may use the aforementioned model as a decision-making tool. With only a small amount of computational labor, the recommended technique may fruitfully recognize the leaf illnesses with an accuracy of 94–95%.

REFERENCES

[1] J. Belasque Jr, M. Gasparoto, and L. G. Marcassa, “Detection of mechanical and disease stresses in citrus plants by fluorescence spectroscopy,” Applied Optics, vol. 47, no. 11, pp. 1922-1926, 2008.

[2] A.-K. Mahlein, “Plant disease detection by imaging sensors-parallels and specific demands for precision agriculture and plant phenotyping,” Plant Disease, vol. 100, no. 2, pp. 241-251, 2016.

[3] Usama Mokhtar et al. “SVM-based detection of tomato leaves diseases”. In: Intelligent Systems’ 2014. Springer, 2015, pp. 641–652.

[4] S. D. Khirade and A. B. Patil. “Plant Disease Detection Using Image Processing”. In: 2015 International Conference on Computing Communication Control and Automation. Feb.2015, pp. 768–771. DOI: 10.1109/ICCUBEA.2015.153.

[5] Y. Tian, P. Zheng and R. Shi, "The Detection System for Greenhouse Tomato Disease Degree Based on Android Platform," in 2016 3rd International Conference on Information Science and Control Engineering (ICISCE), Beijing, 2016

[6] A. K. Hase, P. S. Aher and S. K. Hase, "Detection, categorization and suggestion to cure infected plants of tomato and grapes by using OpenCV framework for andriod environment," in 2nd International Conference for Convergence in Technology (I2CT),Mumbai, 2017.

[7] A. Vibhute and S. K. Bodhe, "Applications of Image Processing in Agriculture: A Survey," International Journal of Computer Applications, vol. 52, no. 2, pp. 34-40, 2012.

[8] J. G. B. Garcia, "Digital Image Processing Techniques for Detecting, Quantifying and Classifying Plant Diseases," SpringerPlus, 2013.

[9] A. M. Mutka and R. S. Bart, "Image-Based Phenotyping of Plant Disease Symptoms," Frontiers in Plant Science, vol. 5, pp. 1-8, 2015.

[10] S. P. Mohanty, D. Hughes and M. Salathe, "Using Deep Learning for Image-Based Plant Disease Detection," eprint arXiv:1604.03169, 2016.

[11]. D. Ashourloo, H. Aghighi, A. A. Matkan, M. R. Mobasheri and A. M. Rad, "An Investigation Into Machine Learning Regression Techniques for the Leaf Rust Disease Detection Using Hyperspectral Measurement," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,vol. 9, no. 9, pp. 4344-4351, 2016.

[12]. Parikh, M. S. Raval, C. Parmar and S. Chaudhary, "Disease Detection and Severity Estimation in Cotton Plant from Unconstrained Images," 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 594-601, 2016.

[13]. H. Sabrol and K. Satish, "Tomato plant disease classification in digital images using classification tree," 2016 International Conference on Communication and Signal Processing (ICCSP), 2016, pp. 1242-1246.

[14] Sharada P Mohanty, David P Hughes, and Marcel Salathe.´ "Using deep learning for image-based plant disease detection".In: Frontiers in plant science 7 (2016), p. 1419.

[15] Christian Szegedy et al. "Going deeper with convolutions". In: Cvpr. 2015.

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton."Imagenet classification with deep convolutional neural networks". In: Advances in neural information processing systems. 2012, pp. 1097–1105.

[17] Jihen Amara, Bassem Bouaziz, Alsayed Algergawy, et al."A Deep Learning-based Approach for Banana Leaf Diseases Classification." In: BTW (Workshops). 2017, pp. 79–88.

[18] Y. Dandawate and R. Kokare, "An automated approach for classification of plant diseases towards development of futuristic decision support system in indian perspective," in Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on.IEEE, 2015, pp. 794-799.

[19] D. G. Tsolakidis, D. I. Kosmopoulos, and G. Papadourakis, "Plant leaf recognition using zemike moments and histogram o f oriented gradients," in Hellenic Conference on Artificial Intelligence. Springer,2014, pp. 406-417.

[20] Agarwal, M., Singh, A., Arjaria, S., Sinha, A., & Gupta, S. (2020). ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network. Procedia Computer Science, 167, 293–301. https://doi.org/10.1016/j.procs.2020.03.225

APPENDIX

We had to deal with a lot of challenges in order to finish the project, the first of which was deciding on our methodological strategy. Since there had not been much work done recently on this range, it was not standard work but more an investigation regarding the based venture. So it's possible that we won't receive many offers of help from anywhere. Another problem was that gathering information was really difficult for us. Datasets for detecting tomato leaves were few. With some self-management, we were able to solve the problem.

TOMATO LEAF DISEASES CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK

ORIGINALITY REPORT

21%	20%	10%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	8%
2	research.vit.ac.in Internet Source	2%
3	Submitted to Jacksonville University Student Paper	2%
4	ijariie.com Internet Source	1%
5	Halil Durmus, Ece Olcay Gunes, Murvet Kirci. "Disease detection on the leaves of the tomato plants by using deep learning", 2017 6th International Conference on Agro-Geoinformatics, 2017 Publication	1%
6	Robert G. de Luna, Elmer P. Dadios, Argel A. Bandala. "Automated Image Capturing System for Deep Learning-based Tomato Plant Leaf Disease Detection and	1%