Agro disease finder system of cauliflowers using convolutional neural network

BY

ISRAFIL

ID: 181-16-255

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computing and Information System

Supervised By

Mr. Abdullah Bin Kasem Bhuiyan

Lecturer

Department of Computing & Information System

Daffodil International University

Co-Supervised By

Mr. Md. Selim Hossain

Senior Lecturer

Department of Computing & Information System

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

December-2021

APPROVAL

This Thesis titled **“Agro Disease Finder System of Cauliflowers Using Convolutional Neural Network”**, Submitted by ISRAFIL, ID No: 181-16-255 to the Department of Computing & Information Systems, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computing & Information Systems and approved as to its style and contents. The presentation has been held on- 13-02-2022.

BOARD OF EXAMINERS

Mr. Md Sarwar Hossain Mollah — **Chairman**
Associate Professor and Head
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Ms. Nayeema Rahman — **Internal Examiner**
Sr. Lecturer
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

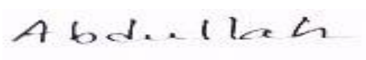
Mr. Abdullah Bin Kasem Bhuiyan — **Internal Examiner**
Lecturer
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Dr. Mohammad Shorif Uddin — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University, Savar, Dhaka

DECLARATION

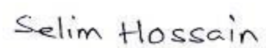
I hereby declare that; this thesis has been done by me under supervision of Mr. Abdullah Bin Kasem Bhuiyan, Lecturer, department of Computing and Information System (CIS) of Daffodil International University. I am also declaring that this thesis or any part of there has never been submitted anywhere else for the award of any educational degree like, B.Sc., M.Sc., Diploma or other qualifications.

Supervised By



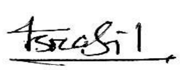
Mr. Abdullah Bin Kasem Bhuiyan
Designation: Lecturer
Department of CIS
Daffodil International University

Co-Supervised by:



Mr. Md. Selim Hossain
Former Senior Lecturer
Department of CIS
Daffodil International University

Submitted By



Name: ISRAFIL
ID: 181-16-255
Department of CIS
Daffodil International University

ACKNOWLEDGEMENT

At first I express my heartiest thanks and gratefulness to almighty God for his divine blessing and makes me possible to complete the final year thesis successfully.

I really grateful and wish my profound my indebtedness to, **Mr. Abdullah Bin Kasem Bhuiyan Lecturer**, Department of Computing & Information System, Daffodil International University, Dhaka and **Mr. Md. Selim Hossain**, **Former Senior Lecturer**, Department of Computing & Information System, Daffodil International University, Dhaka. Their deep knowledge & keen interest with supportive instructions helped me in the fields of Data Analysis, Machine Learning, Deep Learning & Natural Language Processing to carry out this research. Their 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 research.

I would like to express my heartiest gratitude to **Mr. Md Sarwar Hossain Mollah**, **Associate Professor and Head**, Department of Computing & Information System, Daffodil International University, Dhaka, for his valuable support and advice to finish my research and also heartiest thanks to other faculty members and the staffs of Department of Computing & Information System, Daffodil International University, Dhaka.

At last, again I want to thank all the good wishers, friends, family, seniors for all the help and inspirations. This research is a result of hard work and all those inspirations and assistance.

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

ABSTRACT

In an underdeveloped country in South Asia like Bangladesh, every year, many farmers have suffered significant losses as a result of the cauliflowers illness. Farmers have no advanced knowledge of infection detection and mitigation strategy. They can understand the infections after the cauliflowers had already been damaged and in that, they have nothing to do in this matter. Many farmers are now afraid to take efforts to grow cauliflowers because of the loss of production. In this regard, we conducted a study using the improvement of artificial intelligence technology to identify and categorize cauliflowers sickness. We provide an online computer vision technology for developing an Agro disease finder system (AdFS) that analyzes a cauliflower picture recorded with a mobile or portable device and identifies illnesses, allowing faraway farmers to treat the problem. Firstly, we made a dataset with the help of agricultural expertise. They help us to identify the unhealthy cauliflower by examining it in their laboratory after that we took the photo and made this dataset for our research work. Our dataset contains 444 infected cauliflower images which categorize into four types of diseases. According to the TensorFlow and Keras APIs, we utilized the CNN model. This model is dependable and completely linked with all segmentation accomplished. The whole procedure is dependent on deep learning and the technique we use here is called the transfer learning technique. For accomplishing this problem, we used three state-of-the-art algorithms and that is VGG19, VGG16 and ResNet50.

TABLE OF CONTENTS

CONTENTS **PAGE**

CHAPTER **PAGE**

LIST OF FIGURES AND TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Bangladesh is a growing economy according to the World Bank and the size of its GDP is 324.24$ billion. [1]. In this large GDP of Bangladesh, agriculture has a great contribution of around 14.2% [2]. That's why, agriculture has an essential part in the socio-economic development of Bangladesh, such as nutrition, employment, food security, and agricultural exports, and so on. Due to a lack of sufficient training, sophisticated instruments and techniques, education, and other factors, most peasants in Bangladesh do not harvest using scientific methodology. The majority of the time, researchers focus on detecting fruit disease faults. However, the approach is divided into two halves. The first is fault detection, and the second is fault categorization. We can find the best strategy to avoid fruit disease by identifying the defect. Some of the researchers used image processing and mathematical expressions to detect the sickness based on this visual. They would be able to supply more harvests and use their assets more effectively and they have sufficient training, sophisticated tools, and methodologies. We picked cauliflowers from among various agricultural goods since they are commonly cultivated and popular vegetables in Bangladesh. Cauliflower disease causes farmers to lose money since it has a major impact on both the volume and quality of produce. This is a big economic damage for both the farmers and the country. As a result, identifying illness and determining the period of disease infection can be critical to the effective development of cauliflower. In the field of agricultural research, our cauliflower images will be a valuable source of data. In this study, we used a Deep Learning technique to conduct research on cauliflower disease identification. Our approach uses a Convolution Neural Network (CNN) with transfer learning technique to accept cauliflower photos directly, train the model for accuracy testing, and ultimately detect diseased cauliflowers.

1.2 Objective

The agricultural research sector is quickly increasing and evolving, thanks to artificial neural networks. Computational analysis is being used by researchers all around the world to help the agriculture industry thrive. Cauliflower study was chosen since it is a significant contributor to our agriculture. Many species are on the edge of extinction due to a lack of care and support. So, in the future, our objective is to create an intelligent system that will allow us to classify cauliflower diseases into categories and construct an android-based application system to assist our farmers in promptly identifying cauliflower diseases and receiving treatment recommendations. So, we are going to describe some of our goals in points below:

- Our initial objective is to categorize the various types of cauliflower disease.
- Patterns from many categories of cauliflower disease are searched for, and the pattern is taught to the machine through a Convolutional Neural Network.
- In the future, creating an android-based intelligent system that can classify cauliflower disease and immediately propose cures based on the ailment.

1.3 Motivation

In the agricultural sector, image-based disease recognition is a huge field for research, and most of the agricultural goods-related research has not been done yet. We first studied a lot about disease recognition-related papers and then we chose some core topics related to the disease recognition problems. After that, we read some research papers related to those topics. We did study the limitation and future scope of those papers, and finally, we found that some researchers worked on disease classification on agricultural goods. We also saw that already many researchers worked on various agricultural products like tomato, papaya but the cauliflower disease classification research was not found. So, we decided that we will work on the cauliflowers to identify the disease with Convolutional Neural Network.

1.4 Rational of the study

There has never been any research done on the illness categorization of cauliflowers. The majority of the research has focused on disease diagnosis and categorization in various agricultural crops. However, just a little amount of study has been done on the remedies suggestion. Furthermore, only Machine Learning approaches have been used to recognize diseases. So, for the first time, we suggest Deep Learning strategies for identifying cauliflower illness detection, and we compared it to previous research using different Convolutional Neural Network techniques. Previous studies focused on a single condition, but we gathered and created a dataset that included a variety of diseases. As a result, we gathered 444 illness photos with fresh cauliflowers and based our model on them.

1.5 Research Questions

We faced several obstacles before beginning our study because gathering data from the agricultural field is difficult and time-consuming. We also had trouble developing our models and detecting hidden trends in data from the same illness category. So, prior to doing our study, we picked a few questions and concentrated our efforts on answering and concluding those questions.

- Where can we get the data we need for the study?
- Is it feasible to completely pre-process the dataset in order to uncover the underlying pattern of the many types of cauliflower disease?
- Is it feasible to compare the accuracy of different algorithms while working with the same dataset?

1.6 Expected Outcome

As previously said, we employed Deep Learning techniques for the first time to diagnose cauliflower sickness and attempted to create an autonomous Android application for future usage. For the first time in the cauliflower study, we created a dataset.

- It will be able to classify the cauliflower disease.
- To build an agro-disease of cauliflower dataset.
- For the categorization of cauliflower sickness, a novel Deep Learning technique approach is presented.

1.7 Layout of the Report

In Chapter 1 we will discuss about introduction of our research with objective, motivation, research questions and expected outcome.

In Chapter 2 we will discuss about the literature review part with introduction behind that section and challenges on our research.

In Chapter 3 we will discuss about the Agro Disease Finder System (AdFS) architecture. Moreover, system workflow diagram.

In Chapter 4 we will discuss about system methodology, Deep learning approaches. We also discuss about vgg19, vgg16 and resnet50 algorithm.

In Chapter 5 we will discuss about Dataset description, cauliflower Disease Description, Dataset Preprocessing and Feature Extraction.

In Chapter 6 we will discuss about the Result as well as the Discussion. We will also show the graph of the measurement of the model.

In Chapter 7 we will discuss about conclusion and future work of our research. We will briefly describe what are the limitations in our research and how we can overcome those in future and also describes some points related to future aspects of our research.

CHAPTER 2
BACKGROUND STUDY

2.1 Introduction

This section will cover literature reviews, research obstacles, and a comparison of our findings to earlier studies. We will also compare our work to other research papers' approach, works, outcomes, accuracy, and so on. In the Challenges section, we'll explain how we overcame obstacles to conduct our study and what we learnt from them.

2.2 Literature Review

M. T. Habib [3] and his colleagues employed K-means clustering for picture separation and SVM to discover inspection outcomes in their research. Even though they reached a high level of accuracy, they did not execute the test for a huge feature set where we could accomplish it quickly. D. A. Dudwadkar [4], doing research about the corps with some AI based experiment with the object's pH and others. But they didn't summarize the process for image classification, recognition or detection. P. P. Ashish [5], they focus on using the K-nearest neighbor technique to identify and classify plant leaf diseases. However, they did not acquire the predicted result as expected and did not disclose it in their study. R. K. Veeraballi [6], they generally research with papaya curl leaves and use the CNN classification algorithm to diagnose the ailment. They did not, however, deal also with fruits. They have achieved an average of 85% accuracy for detecting leaf disease. S. K. Behera [7], they conducted study utilizing the KNN and SVM classification algorithms to determine papaya ripeness based on its layer. They were unaware of the illness and used machine learning to get an accuracy rate of 90% for papaya ripening status. Pothen [8], they used an image processing approach using a HOG & SVM classifier model to study leaf disease. Based on their technique, they employed HOG to detect the dividing layer. Vijayakumar [9], to determine the fruity sweetness of a fruit, the study used a deep neural network and RESTNET152 to recognize it. Chen, Joy Iong Zong [10], they used a machine learning technique to investigate curvature of harvested crops utilizing the texture degradation option. Kelly [11], they analyzed the impact degree of fruits on the body in this study. For their discoveries, they also researched on temperatures and other bio elements. Liang [12], using CFD and tests, they investigated how to construct a multi-duct cleaning mechanism for rice combine harvesters. Chandy [13], the researcher investigated RGBD analysis of various stages of ripe and immature fruits in this study. Manoharan [14], the authors attempt to do study on hermitical graph wavelets in feature detection in this paper. Later on, the research yielded reliable data that might be used in future studies. I. Hamim [15], they conducted study to get a survey of impacted fruits illness and the disease condition based on previous years' data in their work.

Shiv [16], they found a terrific way for detecting the defect from a batch of picture data by processing mathematical equations and graph visualization throughout their investigation. They do not, however, make much progress in this sector. Bhavani J. Samajpati [17], with the supplied dataset, the K-means clustering technique was used to detect and classify diseases in fruits. However, their study is not particularly practical for non-technical people or farmers who want to identify or diagnose illness in their fruits. They employed a binary value pattern, textures, and other features. Md. Helal sheikh [18], only the utility of defect detection of tree leaves has been tested. They used the Adam method to perform image analysis on a fixed data set in order to reduce the learning rate. As a result, they didn't use any other categorization techniques. Kawaljit Kaur [19], they conducted study just to learn about the condition and procedures in theory. There is no actual work in their projects, and they just supply knowledge that differs from year to year. Fabio Augusto Faria [20], they used a bootstrap aggregator machine learning approach to extract illness information from a sample collection. They designed a program method and employed a support vector machine to obtain value and graph representations depending on their condition. Nikhita M [21], they employed the inception V3 model to separate picture layers and get value from image processing for illness classification. They used a variety of datasets to gather dark or diseased spots on fruit and compute an arithmetic average to produce the desired result. In this study, there are no accuracy measurements. Santi Kumari Behera [22], in their study, they used surface-based image analysis to diagnose illness in fruits. Based on the result from the image collection, they performed pre - processing, feature extraction, and classification. However, no vision-based method or achievement accuracy is displayed. From start to finish, they employed fuzzy logic for interactions. Zechen zheng [23], they use data science and intelligent systems to explore fruit tree disease identification. By analyzing the leaves and fruits, they were able to obtain information from them using convolutional neural networks. They explored with their data by using an algorithm. Despite the fact that they obtained the fault detection value by using the method for the data set. Priyanka soni [24], they conducted study using images of leaves, fruits, or any other type of vegetable, evaluating the frequency of good and poor areas to gain a better understanding. They generalized the data into a data set for illness classification after collecting the details. However, in comparison to the fruits illness, the precision measurement and methodologies are not advanced enough. As a result, we create a robust CNN model using Vgg19 algorithm based on the preceding situation and for our research purposes. The CNN model will retrieve the featured picture first, and then we will use machine learning to apply Vgg19 to acquire the required result.

2.3 Challenges

Cauliflower is a seasonal food, making data collection difficult and complicated at this point in the season. During our investigation, we ran into a variety of issues. First, we had difficulty creating our dataset since it was difficult to search for and locate specific disease cauliflower. As a result, we used agricultural expertise to obtain disease cauliflower data. However, after compiling the data, we discovered that certain categories had relatively little datasets.

We attempted to collect three distinct types of cauliflower data and create a dataset, however, we encountered numerous obstacles in obtaining a better result for a small portion of the data. After collecting data, we worked on data pre-processing and faced problems to give a particular shape. The hardest task was to clean the dataset. But we enjoyed our challenges and learned from them.

CHAPTER 3
SYSTEM ARCHITECHTURE

3.1 Introduction

In this chapter, we will discuss the system workflow and its working step. Besides this, we will discuss about how AdFS accurately identify the cauliflower disease and classify them into their categories.

3.2 System Workflow

The architecture of "Agro-disease finder systems (AdFS)" of cauliflowers for identification the cauliflowers disease is given in Fig 1.

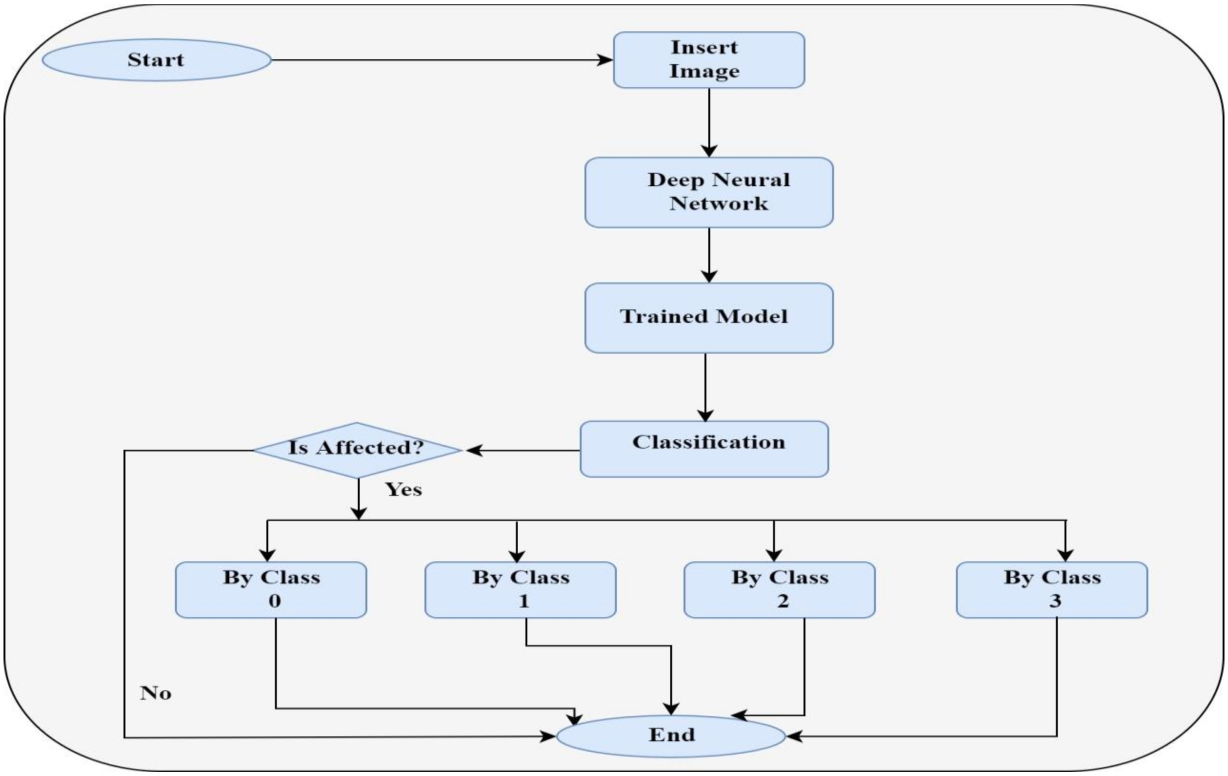


Fig. 1: Workflow diagram for Agro-disease finder systems (AdFS)

This system uses a Convolutional Neural Network to anticipate the outcome. A farmer takes photographs of the cauliflower so that the unhealthy of the cauliflower may be easily determined. If it is infected with a disease then it will notify the farmer via text-box message. Usually, it will be an android-based application that farmers can examine in a minimal period of time. Our farmers just need a smartphone and internet connectivity to do their tasks. After uploading a picture, the Keras API is used to retrieve image characteristics in the rear and to forecast the outcome applying logistics regression. In the front-end, the user may see the result via text-box message.

The workflow of cauliflowers illness categorization that we developed is shown in Fig 1. To begin, we must first create an image data or database of size (224×224×3) at the beginning part. Finally, it moves on to the hidden layer, where the Convolution layer, MaxPooling, and ReLU functions are used to choose data features, and then the classification portion uses Flatten, Dense, and sigmoid functions to classify the model. Following categorization, the suggested model predicts which type of disease of that particular cauliflower has or predicts it is a fresh cauliflower which is what we most desire.

CHAPTER 4

SYSTEM METHODOLOGY

4.1 Introduction

We created a full roadmap of techniques to identify disease-affected cauliflower and fresh cauliflower, and we are using Convolution Neural Network (CNN) to identify the cauliflower dataset, with transfer learning technique being the most commonly used.

4.2 System Methodology

The suggested method is unique in that it makes use of a low-cost, high-accuracy system that is defined as a deep learning model. Similar machine learning algorithms are slower than the CNN (Convolutional Neural Network). Using the transfer learning technique, it can quickly diagnose illnesses.

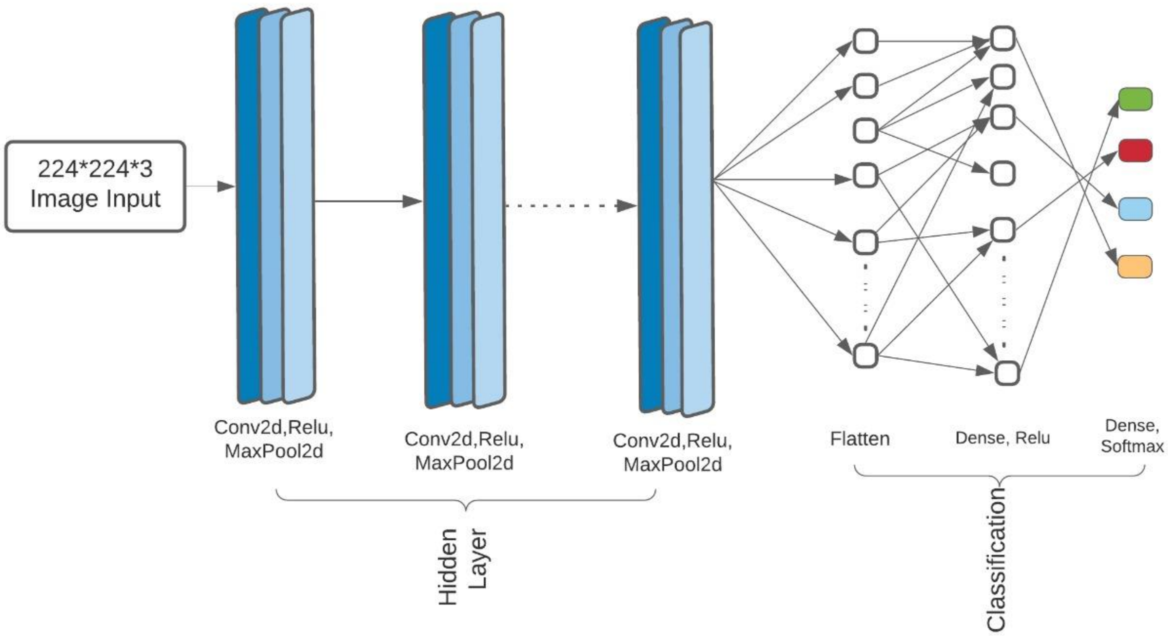


Fig. 2: Classification of disease of cauliflower using convolutional neural network (CNN)

We mostly categorize diseases using RGB images with a 224×224 input form. Then, inside the hidden layer, the transferable convolutional layer offers an outcome shape of (222,222,16), while the non-learnable layer MaxPool simply divides the output form by 2 and produces a new output form of (111,111,16). It is also used to lower the transferable layer as well as the MaxPooling layer by half using the same technique. This computation is accomplished using the following rule:

$$x_h = x_w = \frac{x + 2p - f}{s} + 1$$

As in CNN model, this would be the formula for output changes. For input picture data, x_h stands for "image height" and x_w stands for "image weight," and both are predefined for our models. Padding (p) is a concept related to CNN that represents the number of pixels on an input picture in a CNN model, and padding is 0 in the suggested model, implying that the extra pixel intensity is 0. Here kernel shape is (f) which defines with the batch size by 3. So, stride is the count of pixels that shuffle across the weight matrix, and we've set that to 1, which means it can only filter to one pixel at a moment.

After the hidden layer, this is the prime part for image classification of both fresh cauliflower and sick cauliflower and it is called the "Classification" section. For the transferable layer, it modifies the parameters, and below in the Fig. 3 shows the model summary as to how our model modifies its parameters.

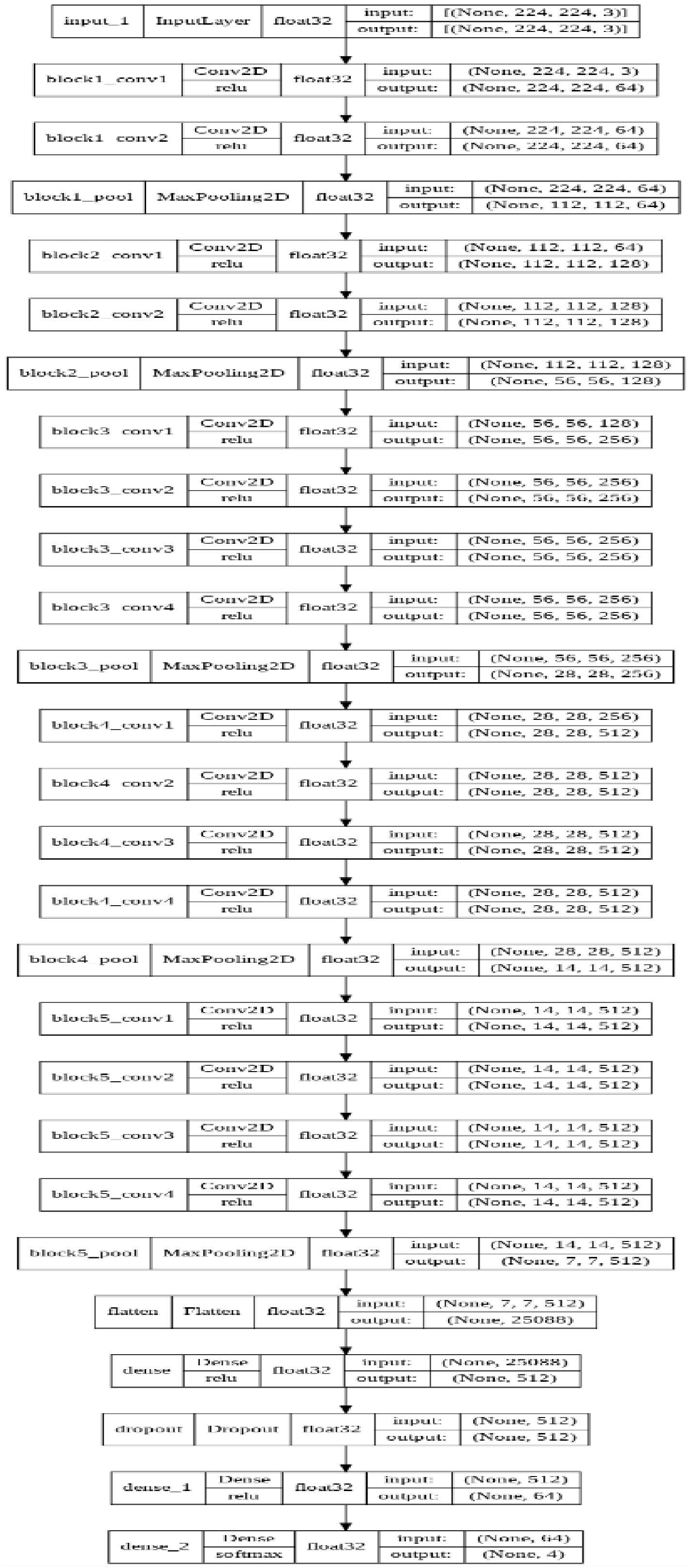


Fig. 3: Model summery of cauliflower disease recognition

Now, as shown in Fig. 3, the last output form of the Convolution Layer and MaxPooling layer has been flattened, and even the Dense layer has completed its task, and it has been effectively categorized and is also called as output. We had used OpenCV to evaluate our system so that it could accurately predict whether the illness affects or fresh cauliflower.

4.3 Deep Learning Approach

We applied Deep Learning techniques for the first time to classify cauliflower disease recognition systems into meaningful categories. We'll go through how we employed Deep Learning models to solve this challenge, as well as the algorithm we used. VGG19, VGG16, and ResNet50 are three state-of-the-art algorithms that we use. Here we adopt a vital technique called the transfer learning which is also discussed below.

Transfer Learning Technique

In simple terms, transfer learning is leveraging others' pre-trained models for our task; the task can be different, but it must be connected. Deep learning models have been shown to be transferrable to some extent in the literature.

We looked into two methods for transfer learning: using the pre-trained model as feature extraction or retraining (fine-tuning) a portion of the model. It may be applied quickly and easily while maintaining the model's original weight. The fine-tuning procedure, on the other hand, necessitates some trial and error. If the jobs are substantially diverse, we may want to fine-tune at least half of the model, if not all of it. If our jobs are quite comparable, for example, if the model was originally used to categorize flowers and we now want to diagnose the cauliflower disease, we could fine-tune the last few layers or up to 30% of the original model. Depending on the nature of the activity, the amount of fine-tuning required takes time and effort to investigate.

There are several recognized models for disease recognition tasks, and we have covered three of them with our example for this research.

VGG19 Algorithm

The VGG-19 has been the first model we looked at and the oldest of the models we looked at. The VGG-19 is an upgrade of the VGG-16 model. It's a 19-layer convolutional neural network model. It is constructed by stacking convolutions together, however, the depth of the model is restricted due to a phenomenon known as decreasing gradient. Deep convolutional networks are difficult to train because of this problem. The model, like the rest of the models examined, was trained on ImageNet to classify 1000 different categories of objects.

VGG-19 was investigated as a feature extraction and fine-tuning tool. The best result we've gotten so far is from utilizing VGG-19 as a feature extraction tool. In our experience, fine-tuning and retraining do not function well with VGG-19.

VGG16 Algorithm

Edge and color information are commonly seen in the features of the earlier layers of a pre-trained CNN. The subsequent layers, on the other hand, have properties that are more particular to the classes' features. As a result, the parameters of the previous layers require no or little fine-tuning. As a result, only the last three layers of the VGG-16 have been fine-tuned in this study.

The following is an explanation of how to change the layers. The VGG-16 has been trained on over a million photos. It can classify photos into 1000 different categories. These 1000 classes are configured in the VGG-16's last three tiers. These will need to be finetuned for a newer categorization task. Fine-tuning is accomplished by extracting all of the network's layers except the last three. Substitute a fully connected layer, a softmax layer, and a classification output layer for the last three levels to shift the layers to the new classification job. The completely linked layer's size is set to the same as the number of classes in the new data.

The size value for our work is four, which corresponds to the number of classes: Alternaria Leaf Spot, Cabbage aphid colony, clubroot, and ringspot. The whole process for classifying cauliflower diseases utilizing the VGG-16 network's transfer learning mechanism.

Resnet50 Algorithm

ResNet is the abbreviation for Residual Network, which is seen in Figure 7. Deep convolutional neural networks have produced a number of improvements in image identification and classification over the years. It's becoming fashionable to dig a bit deeper to answer more difficult problems and increase classification or recognition accuracy. However, due to issues such as the vanishing gradient problem and degradation problem, training deeper neural networks has proven problematic. Residual learning aims to address both of these issues.

While being trained for the job at hand, each layer of a neural network learns low or high-level characteristics. Instead of trying to learn features, the model tries to learn some residual in residual learning. ResNet50 is a 50-layer Residual network with ResNet101 and ResNet152 as variations. The results of using ResNet as a pretrained model for cauliflower disease categorization have been promising.

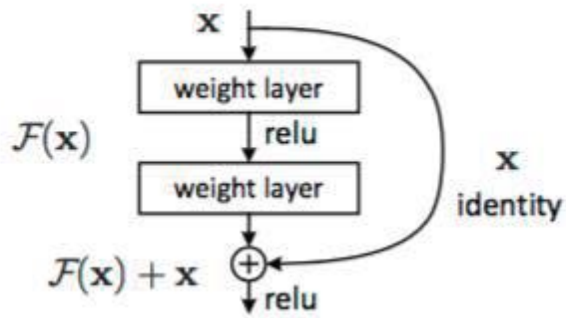


Fig. 4: Residual learning: a building block

A RGB (Red Green Blue) picture is used as the model's input. The picture will be fed into the ResNet50 layer, which will include the pretrained weights, and the final layer will be a fully linked dense layer with sigmoid activation. The proposed model has two layers: a dense layer and a pre-trained Resnet layer. Weights that have been pre-trained for the ResNet50 model will be imported. The dense layer is the sole layer that learns via back propagation, and the input data will be taught with pre-trained weights. Few levels, such as Batch Normalization (BN) layers, should not be frozen since the dataset's mean and variance will be quite different from the pre-trained weights' mean and variance.

CHAPTER 5
DATASET & DISEASE DESCRIPTION

5.1 Dataset Description

444 images were used to forecast the approaching model, 426 images were used to the training set, and 18 images were used to test it. We made our dataset on ourselves with the help of agricultural expertise. Besides this, we collected some cauliflower disease-related images from the internet. Our dataset name is Agro-disease finder of cauliflower and the dataset location is (https://www.kaggle.com/shahriarahmedisrafil/dataset-of-disease-cauliflower). Certain disease forms such as Alternaria Leaf Spot, Cabbage aphid colony, clubroot, and ring spot are included in the disease data sets. The research dataset and the method dataset are not identical; to improve the accuracy of the Deep learning model, we are expanding the quantity of real picture data. The height, width, and RGB color code of the input picture data are employed in the Deep Learning model as three dimensions. First, all photos were transformed to 224×224×3 pixels in our model. As a result, the Deep Learning model primarily works with the RGB color of a cauliflower picture, which is then classified.

5.2 Disease Description

This portion is a critical component of our study strategy since it allows us to correctly analyze the disease, comprehend the fault, and learn more about appropriate features. We classify the four types of cauliflower disease to recognize our model. These four illnesses are quite widespread in Bangladesh. The majority of farmers are dealing with the following illnesses in their crops: Alternaria Leaf Spot, Cabbage aphid colony, clubroot, and ring spot are all examples of cauliflower disease.

Alternaria Leaf Spot

The fungal pathogen Alternaria brassicicola causes Alternaria leaf spot, a widespread foliar disease of brassica crops. Many brassica crops, such as cabbage, cauliflower, kale, brussels sprouts, and broccoli, are susceptible to the illness. Even minor diseases might result in a harvest that is unsellable. Severe foliar infections might result in a drop in yield owing to leaf loss or weight loss.

Fig. 5: Affected by Alternaria Leaf Spot

Pin-sized black specks emerge on the leaf surface or branch as the first symptom. The specks grow in size in a concentrical pattern, forming a recognizable target spot or bull's eye lesion with a yellow halo. The necrotic lesions can readily rip and are evident from both sides of the leaf. Infections on the leaves can cause the leaves to fall off the plant. During prolonged periods of high humidity, black sooty spores form on the surface. In addition to the leaves, symptoms might emerge on cauliflower curds or broccoli heads. Symptoms usually appear first on older, lower leaves.

Cabbage aphid colony

Cabbage aphids have a white waxy covering on their bodies and are green-gray in color. They live in thick colonies and are frequently coated in waxy droplets. They love the young leaves and blooming sections of cabbages and Brussels sprouts and are frequently found deep within the heads of these vegetables. Adult females give birth to live progeny throughout the year, making the aphid's life cycle uncomplicated. Adults with wings and adults without wings can be found; the winged adults have a black thorax and lack a waxy covering. The aphid does not infest non cruciferous crops, but it can live in the absence of cole crops by feeding on allied weed species.

Fig. 6: Affected by Cabbage aphid Colony

Cabbage aphids don't usually bother seedlings, but they might become a problem after thinning or transplanting. Large colonies can stunt or kill young plants, but contamination of the harvested crop is the most significant issue. When populations are dense, leaves coil around them, making pesticide applications more difficult.

Clubroot

Clubroot is caused by a fungus-like microbe that may live in soil for several years after an infected crop causes the illness. Clubroot is thought to infect all members of the cruciferous family, including wild mustard, although cauliflower is one of the most vulnerable. Within the clubroot pathogen population, there are races, and different strains of clubroot infect different cruciferous species.

Fig. 7: Affected by Clubroot

Atypically big roots-fine roots, secondary roots, the taproot, or even on the subterranean stem-are the defining symptom. When sensitive plants are infected when still in the seedling stage, they can die, and infections at this point can have a big influence on yields since the seedlings that survive are severely stunted. The disease seldom kills plants that are attacked later in their life cycle, although severely deformed roots have a reduced capacity to absorb nutrients and water from the soil. In warmer temperatures, plants wilt but recover somewhat at night.

Ring spot

Ring spot, sometimes known as black blight, is spread by the ascomycete Mycosphaerella brassicicola, a kind of sac fungus.

Fungal colonization may be ongoing if the foliage has brownish lesions with yellowish edges and a bull's-eye look. If left unchecked, the imperfections will coalesce and turn the leaves black.

Fig. 8: Affected by Ring Spot

Symptoms might appear across the plant's whole surface. Individual lesions contain dark concentric rings with defined boundaries surrounded by a yellowish zone, formed by black fruiting bodies. Spots congregate under severe assault, and the entire plant may be harmed and darkened.

5.3 Dataset Preprocessing

In our deep learning-based cauliflower categorization, we employ a variety of various forms of picture data. We mentioned some image data sizes here that we preprocess for our model and that is 720×350, 400×250, 850×520(from test dataset, in pixel), 1920×1080, 1024×768, 1440×1080, 200×200(from training image data, in pixel). As a result, preparing an image dataset for classifying without a defined size of picture data is too complex. So, in our deep neural network, we utilized the input shape functions with a specified size of 224×224 pixels and an RGB picture dataset, and our model train is very fast and properly classifies the illness to use this input shape.

5.4 Feature Extraction Process

There are various separate levels in this strategy that are useful for transmitting input and output data. Firstly, the Convolutional Neural Network must use a 224×224×3 input picture and set the input layer to the final max-pooling layer (7×7×512). This layer is used for extracting features in the approaches model, while the remainder of the network is used for classification. In the CNN model, the hidden layer and classification parts have been finished, and the hidden layer is mostly utilized for processing image input and feature extraction. Then, using our model, a flattened and dense layer is used to classify it as disease-affected or fresh cauliflower. After conducting our CNN model, we must fit our system to the CNN model while using our own dataset as the training dataset. Then, in comparison to the training model, we forecast our model by using a test dataset.

CHAPTER 6
RESULT ANALYSIS & DISCUSSION

6.1 Discussion

We used the VGG19 algorithm to create a convolution neural network (CNN) model in this study. To begin, we gathered a data set on cauliflower infections and processed it by reducing the shape and altering it. Cauliflower infection photo identification and prediction were done in this study. We divide our image dataset into two parts one is the training part which will train our model and another one is the validation part which will test the model result. In this research, the training and testing ratio keeping 80:20. Keras conv2d is a fundamental 2D layer for this system, and it generates a convolution kernel that is wound with layers input, resulting in a tensor of outcomes. We can see a diseased cauliflower in the image above. Our CNN architecture, which uses the VGG19 algorithm, was able to categorize the picture layout and use a neural network to discover the patterns of the impacted region. We next tested the remaining dataset and found that it had a prediction performance of 84%. In which our model performs better than existing other agricultural disease identification models.

We restructure (224, 224, 3) the picture during pre-processing for cauliflowers leaf categorization and recognition. We next use VGG19 algorithms to conduct a CNN neural network to discover the confusion matrix of the training dataset. Lastly, we checked the model's correctness by fitting the training dataset to it.

6.2 Results

Many of our scientists have used numerous algorithms to identify other agricultural disease recognition for the benefit of farmers, such as K-means clustering and random forests, etc. However, by combining the CNN neural network with the help of the VGG19 algorithm, we were able to outperform the competition and achieve an average accuracy of 84 percent. This study algorithm aids in the more accurate diagnosis of cauliflowers disease.

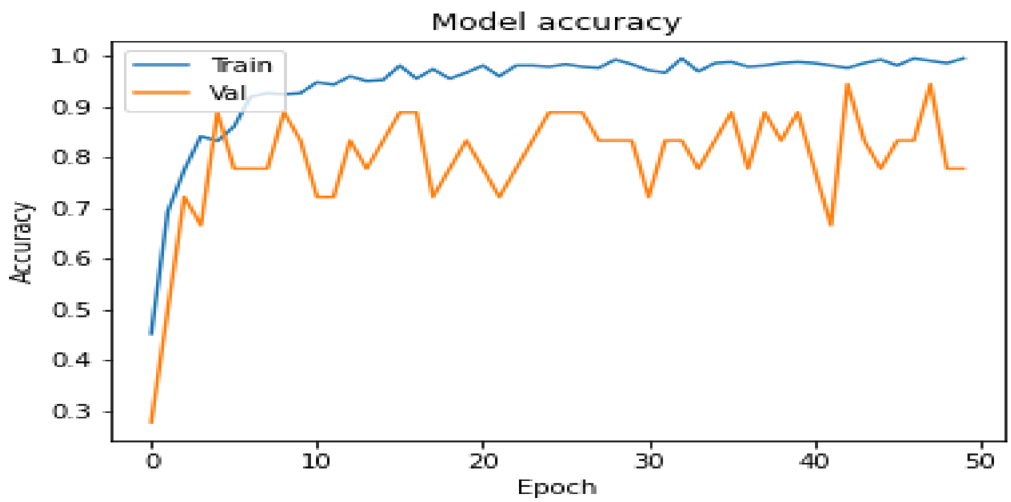


Fig. 9: Accuracy graph vs Epoch of the VGG19 model

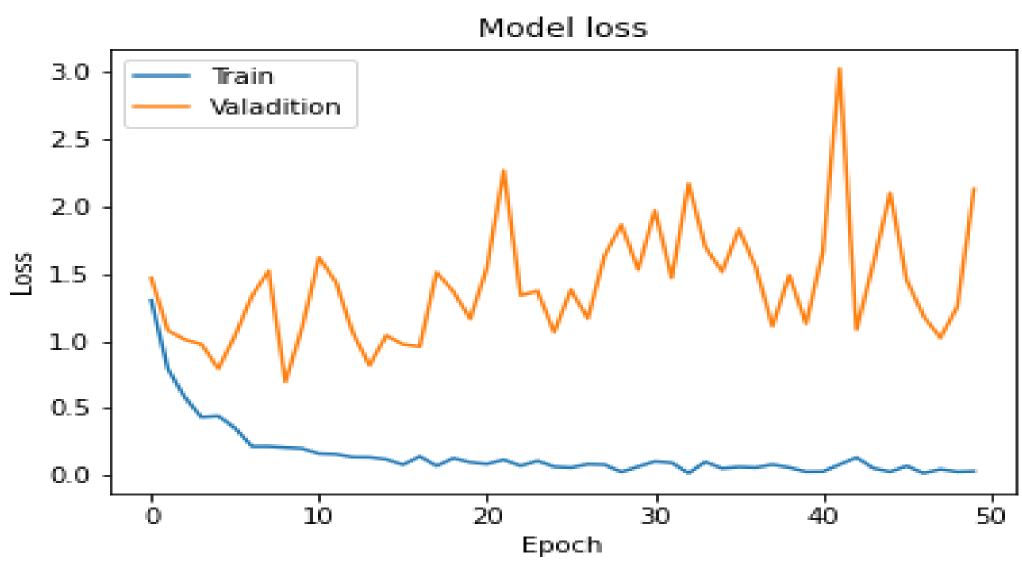


Fig. 10: Loss graph vs Epoch of the VGG19 model

Our accuracy level and loss graph for the training and test datasets are displayed in the diagram above. Our predictive algorithm provides an overall accuracy of 84 percent.

Table. 1. Comparing of different model of cauliflower disease identifier systems

SL No	Model Name	Validation Accuracy	Validation Loss
1	VGG19	84%	1.87
2	VGG16	83%	2.30
3	ResNet50	61%	1.65

From the above table it has been seen that that we have applied different algorithm of this dataset. We have found accuracy of 84 percent by using VGG19 algorithm.

CHAPTER 7
CONCLUSION & FUTURE WORK

7.1 Conclusion & future works

The primary goal of providing this concept is to benefit village farmers and farm owners who do not receive recommendations from the agricultural office. So that they may correctly detect cauliflower disease and obtain the intended production with adequate treatment. Furthermore, by employing this model, the farmer would save time and will be able to precisely identify whether or not the cauliflower is harmed. The major goal of our study is to assist those farmers in obtaining information on their cauliflowers' status using our model and taking the appropriate procedures. Using an image dataset as testing data, our Convolutional Neural Network model predicts if cauliflower is diseased or not. There is a secret aim behind the development of our system and that is to assist farmers all around the globe. This model may be used by not only qualified farmers but also ignorant and distant region farmers to anticipate cauliflower illness as quickly as possible so that they can make the best decision possible to defend cauliflower from disease using a real-time cellphone camera. It will be capable to use tensor flow to extract the cauliflower image as input from video and use it as real-time object detection. It will function by generating discrete segments in the input picture, extracting features, and then converting the bounding region into a unified bounding square to feed into our CNN model which will then estimate if the cauliflower is diseased or fresh. We can improve the accuracy of our model by more than 84% by increasing the number of sick cauliflower photos in our dataset. Following our model's diagnosis of cauliflower illness, we're attempting to recommend a remedy for the ailment to farmers via the Android application.

REFERENCES

[1] J. Kundu, S. Debi, S. Ahmed and S. Halder, "Smart e-agriculture monitoring system: Case study of Bangladesh," 2017 4th International Conference on Advances in Electrical Engineering (ICAEE), 2017, pp. 641-646, doi: 10.1109/ICAEE.2017.8255435.

[2] Tradingeconomics.com. 2021. *Bangladesh GDP | 2021 Data | 2022 Forecast | 1960-2020 Historical | Chart | News*. [online] Available at: <https://tradingeconomics.com/bangladesh/gdp> [Accessed 16 December 2021].

[3] M. T. Habib, A. Majumder, A. Z. M. Jakaria, M. Akter, M. S. Uddin, and F. Ahmed, "Machine vision-based papaya disease recognition," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 32, no. 3, pp. 300–309, 2020.

[4] D. A. Dudwadkar and Vivekanand Education Society's Institute of Technology, "Automated hydroponics with remote monitoring and control using IoT," *Int. J. Eng. Res. Technol. (Ahmedabad)*, vol. V9, no. 06, 2020.

[5] P. P. Ashish and P. Tanuja, "Survey on detection and classification of plant leaf disease in agriculture environment," *Int. adv. res. j. sci. eng. technol.*, vol. 4, no. 4, pp. 137–139, 2017.

[6] R. K. Veeraballi, M. S. Nagugari, C. S. R. Annavarapu, and E. V. Gownipuram, "Deep learning based approach for classification and detection of papaya leaf diseases," in *Advances in Intelligent Systems and Computing*, Cham: Springer International Publishing, 2020, pp. 291–302.

[7] S. K. Behera, A. K. Rath, and P. K. Sethy, "Maturity status classification of papaya fruits based on machine learning and transfer learning approach," Inf. Process. Agric., 2020.

[8] M. E. Pothen and D. M. L. Pai, "Detection of rice leaf diseases using image processing," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 2020.

[9] D. Vijayakumar and M. Vinothkanna, "Mellowness detection of dragon fruit using deep learning strategy," March 2020, vol. 2, no. 1, pp. 35–43, 2020.

[10] D. J. I. Zong Chen and L.-T. Yeh, "Analysis of the impact of mechanical deformation on strawberries harvested from the farm," September 2020, vol. 2, no. 3, pp. 166–172, 2020.

[11] K. Kelly, R. Madden, J. P. Emond, and M. C. do Nascimento Nunes, “A novel approach to determine the impact level of each step along the supply chain on strawberry quality,” Postharvest Biol. Technol., vol. 147, pp. 78–88, 2019.

[12] Z. Liang, L. Xu, J. De Baerdemaeker, Y. Li, and W. Saeys, “Optimisation of a multi-duct cleaning device for rice combine harvesters utilising CFD and experiments,” Biosyst. Eng., vol. 190, pp. 25–40, 2020.

[13] A. Chandy, “Rgbd analysis for finding the different stages of maturity of fruits in farming,” December 2019, vol. 1, no. 02, 2019.

[14] S. Manoharan Dr, “Study on Hermitian graph wavelets in feature detection,” September 2019, vol. 2019, no. 1, pp. 24–32, 2019.

[15] I. Hamim, M. Z. Alam, M. A. Ali, and M. Ashrafuzzaman, “Incidence of post-harvest fungal diseases of ripe papaya in Mymensingh,” J. Bangladesh Agric. Univ., vol. 12, no. 1, pp. 25–28, 2014.

[16] S. R. Dubey and A. S. Jalal, “Application of image processing in fruit and vegetable analysis: A review,” J. Intell. Syst., vol. 24, no. 4, pp. 405–424, 2015.

[17] B. J. Samajpati and S. D. Degadwala, “Hybrid approach for apple fruit diseases detection and classification using random forest classifier,” in 2016 International Conference on Communication and Signal Processing (ICCSP), 2016.

[18] M. H. Sheikh, T. T. Mim, M. S. Reza, and M. H. Hena, “Leaf diseases detection for commercial cultivation of obsolete fruit in Bangladesh using image processing system,” in 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), 2019.

[19] C. M. Kawaljit kaur, “‘Analysis of Diseases in Fruits using Image Processing Techniques,’” in International Conference on Trends in Electronics and Informatics ICEI 2017, 978-1-5090-4257-9/17/$31.00 ©2017 IEEE.

[20] F. A. Faria, J. A. dos Santos, A. Rocha, and R. da S. Torres, “Automatic Classifier Fusion for Produce Recognition,” in 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images, 2012.

[21] Nikhitha, R. Sri, and U. Maheswari, “Fruit recognition and grade of disease detection using inception V3 model,” in 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2019.

[22] S. K. Behera, L. Jena, A. K. Rath, and P. K. Sethy, “Disease classification and grading of orange using machine learning and fuzzy logic,” in 2018 International Conference on Communication and Signal Processing (ICCSP), 2018.

[23] Z. Zheng, S. Pan, and Y. Zhang, “Fruit tree disease recognition based on convolutional neural networks,” in 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS), 2019.

[24] P. Soni and R. Chahar, “A segmentation improved robust PNN model for disease identification in different leaf images,” in 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), 2016.

[25] M. A. Islam, M. S. Islam, M. S. Hossen, M. U. Emon, M. S. Keya, and A. Habib, “Machine learning based image classification of papaya disease recognition,” in 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020.

[26] S. R. Dubey and A. S. Jalal, “Application of image processing in fruit and vegetable analysis: A review,” J. Intell. Syst., vol. 24, no. 4, pp. 405–424, 2015.

181-16-255 Israfil

ORIGINALITY REPORT

16%	6%	9%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Md. Sagar Hossen, Imdadul Haque, Md. Saif Islam, Md. Tanvir Ahmed, Md. Jannati Nime, Md. Ashiqul Islam. "Deep Learning based Classification of Papaya Disease Recognition", 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020 Publication	4%
2	www.jetir.org Internet Source	2%
3	Submitted to Daffodil International University Student Paper	1%
4	Mohammed Abduljabbar Ali, Abir Jaafar Hussain, Ahmed T. Sadiq. "Deep Learning Algorithms for Human Fighting Action Recognition", International Journal of Online and Biomedical Engineering (iJOE), 2022 Publication	1%
5	Submitted to University of Queensland Student Paper	1%
6	Submitted to Indiana University Student Paper	1%

7	Submitted to Teachers' Colleges of Jamaica Student Paper	1 %
8	cpb-us-e1.wpmucdn.com Internet Source	1 %
9	www.mathworks.com Internet Source	<1 %
10	Euphrem Mugisha Rwagasore, Xiong Zhang, Kaifang Gao, Zuoxuan Gao et al. "Convolutional neural network for sapphire ingots defect detection and classification", Optical Materials, 2021 Publication	<1 %
11	www.ijitee.org Internet Source	<1 %
12	www.mdpi.com Internet Source	<1 %
13	Submitted to Ghana Technology University College Student Paper	<1 %
14	"Sustainable Communication Networks and Application", Springer Science and Business Media LLC, 2022 Publication	<1 %
15	Faria Ferdowsy, Kazi Samsul Alam Rahi, Md. Ismail Jabiullah, Md. Tarek Habib. "A machine	<1 %

learning approach for obesity risk prediction", Current Research in Behavioral Sciences, 2021
Publication

16	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1 %
17	hdl.handle.net Internet Source	<1 %
18	www.lib.kobe-u.ac.jp Internet Source	<1 %
19	"Advanced Informatics for Computing Research", Springer Science and Business Media LLC, 2021 Publication	<1 %
20	S Malathy, R.R Karthiga, K Swetha, G Preethi. "Disease Detection in Fruits using Image Processing", 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021 Publication	<1 %
21	dokumen.pub Internet Source	<1 %
22	"Ubiquitous Intelligent Systems", Springer Science and Business Media LLC, 2022 Publication	<1 %
23	R. Manavalan. "Automatic identification of diseases in grains crops through	<1 %

computational approaches: A review",
Computers and Electronics in Agriculture,
2020
Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		