TOMATO PEST DETECTION USING CONVOLUTIONAL NEURAL NETWORK IN BANGLADESH

BY

Johora Akter Polin
ID: 213-25-063

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

Supervised By

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

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2023

APPROVAL

This Thesis titled **"TOMATO PEST DETECTION USING CONVOLUTIONAL NEURAL NETWORK IN BANGLADESH"**, submitted by Johora Akter Polin, ID No: 213-25-063 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 17-01-2023.

BOARD OF EXAMINERS

Chairman

Dr. Sheak Rashed Haider Noori, PhD
Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Ms. Naznin Sultana
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

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

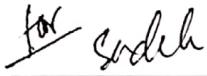
External Examiner

Dr. Mohammad Shorif Uddin, PhD
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

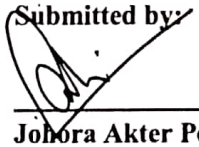
We hereby declare that, this project has been done by us under the supervision of **Dr. Md. Tarek Habib, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:



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

Submitted by:



Johora Akter Polin
ID: 213-25-063
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We really grateful and wish our profound our indebtedness to **Dr. Md. Tarek Habib**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Convolutional Neural Network and Machine Learing" to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this work.

We would like to express our heartiest gratitude to Vice Chancellor, Dean, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

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

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

One of the major commercial crops, tomatoes provide a lot of vitamins and can also be consumed as fruit. Throughout its lifecycle, the tomato is affected by a number of diseases and pests. Lack of prompt management might result in decreased yields or possibly crop loss. The most significant stage in finding out how to effectively control diseases and pests and support vegetable farmers in enhancing tomato production is to accurately identify the diseases and insect pests. The analysis and classification of plant diseases are currently the focus of a wide range of research studies based on image processing. These technologies are useful for rapidly detecting pests and illnesses in plants. Plant pests are still a significant issue for the agricultural sector. Aphids, whiteflies, thrips, red spider mites, and looper caterpillars are just a few of the pests that harm tomato plants. Faster detection of these pests on tomato plants might lead to early treatment and dramatically reduced financial losses. Five different types of tomato bugs were investigated. In this study, we compared two approaches to locating typical tomato bugs. In the first method, CNN is used, whereas, in the second method, CNN is combined with a random forest classifier. The contrast has been provided. We have also used some classifiers to measure the accuracy. This study will serve as a guide for the engineering application of intelligent disease and pest detection.

TABLE OF CONTENTS

LIST OF Tables

LIST OF FIGURES

CHAPTER 1
Introduction

1.1 Introduction

In Asian nations, seventy-five share of the population relies upon agriculture. Bangladesh is an agricultural country. The tomato is one of the most frequent and integral plant life in the world. It's manufacturing in Bangladesh was once estimated to be around 74000 million heaps in 2021, then again dynamic consumption expansion discovered in China and the Republic of India has resulted in plenty of higher global demand. Nowadays, this place is going through one or two problems, like a scarcity of labor or ailments that assault crops, which successively reduce productivity. Therefore, the detection of pests on vegetation at the first stage is quintessential in phrases of heading off giant quantities of pesticides and attaining greater productivity.

Farmers have a huge range of options for choosing different suited crops and identifying the right plant illness. Fungi, bacteria, and viruses spread to the crop due to environmental changes in variables like rainfall, temperature, soil quality, etc. The most vulnerable section of the plant, the leaves, are where disease symptoms first appear. The continuous development of the financial system and society has precipitated worldwide local weather and environmental issues. The incidence of ailments and pests severely affects people's lives[1]. The incidence and occurrence of plant illnesses and trojan horse pests are higher, better, and extra advanced. Therefore, it is integral to take a look at the trouble of plant ailments and malicious program pests, in addition due to the fact of the identification and remedial measures of plant illnesses and bug pests. Tomato is taken into account as one of the quintessential meal commodities.

It's significant everywhere, in every type of environment and soil. The tomato is prone to degradation on a physical, chemical, and microbiological level since it is a very seasonal commodity. Tomato is used for preparing sauces, soup, and many other things[2]. The preparation of the item will take place at temperatures between 65 and 85 degrees C, using either the cold break or the fresh break technique.

The tomato is a significant commercial crop that may be consumed as a fruit in addition to being loaded with vitamins. With the growing popularity of Western cuisine in recent years, spaghetti sauce has become very common. Tomatoes are getting more and more in demand, and they are gradually replacing other staple foods in human diets[3, 4]. As a result, tomato is crucial to the production and commerce of agricultural vegetables. The tomato, one of the most extensively grown vegetables in the world, has a large yield, wide availability, as well as a high biological processing cost. But like other crops, tomatoes are susceptible to several illnesses and pests that affect their growth. The diseases that affect tomatoes include those caused by viruses, nematodes, worms, diseases, physiological illnesses, microorganisms, and fungi. Pests include Polyphagotarsonemus latus, leaf jack, whiteflies, Alfalfa noctuid, tobacco green worm, vetchworm, and other species. The native environment, variety, culture, and management variables often have a significant impact on the incidence of tomato illnesses and pests, and the types of tomato diseases and pests vary substantially among locations.

Regarding tomato categorization, researchers have suggested several practical strategies assisted by computer vision and machine learning. The mission of machine vision involves the acquisition, processing, and analysis of images, making it an effective method for the activity of external factors like color, form, size, and flaws. Together with neural networks and mathematical logic, soft computing techniques—supported by machine vision and previously utilized for pattern recognition and higher cognitive processes—have demonstrated smart results in internal control[10]. Customers today want high-quality produce and meals. A significant problem for the agriculture industry may be fruit grading to determine quality and worth. The standard method for determining tomato quality includes professional inspectors visually examining the most recent fruit; this grading backed by color and appearance will lead to erratic results and reduce the accuracy of fruit categorization.

With three million hectares cultivated globally each year and the production of over sixty million tons of perfect tomato fruit, the tomato plays a significant role in the global agricultural economy. The quality of tomatoes is determined by examining their color and surface features using various automatic categorization equipment. Area units are used to ensuring that the tomatoes and their products are of a high standard[12],[13]. An entire batch of fruit will be damaged by a few broken tomatoes in a box during the following storage and processing.

1.2 Motivation

The purpose of the study was to detect pests using a model of CNN that will be using various computer science techniques to assist farmers, businesses, and consumers with tomato categorization and fault prediction. The study's major objective was to compare deep auto-encoders with other methods. A substitute intelligent approach for tomato fault designation is thus being designed. This might help with productivity and ranking within the tomato food industry.

1.3 Rationale of the Study

The second most often cultivated and consumed vegetable in Bangladesh is tomatoes. Tomatoes often undergo remarkable changes in color, flavor, texture, and composition as they mature gradually from the inside out. The bulk of tomatoes that are consumed is thermally processed into a variety of products, including whole canned tomatoes and components. Numerous quality factors, such as the tomatoes' acidity and softness, influence the choice of fruit for canned whole tomatoes. Handling to feel the feel within the food trade is often used to identify soft tomatoes, which don't look suited for canned meals. Soft tomatoes will be transported to the ingredient assembly line rather than the canned tomato line after they are identified. The cytomembrane's chemical contents may lose their state or degrade, which might lead to tomato softening. The breaking of tomato cell walls, and therefore its structural components, has a pleasant effect on how a tomato feels.

The online detection of the softness and microbiological deterioration of the whole tomato is still fraught with formidable difficulties. A quick technique to increase the effectiveness of tomato quality analysis is desired by buyers, processors, and scientists who focus on tomato breeding and internal control methods. For tomato processors, having a mechanism to swiftly and automatically assess the quality of tomatoes is essential for determining an acquisition's worth [15]. The mechanical texture instrument, nonetheless, isn't the ideal solution for online whole fruit softness detection because of its hazardous nature and low efficacy. From the field to the table, tomato mold infection will happen at completely different periods. These diseased fruits are thrown away as rubbish or used as an organic soil fertilizers. In prior work, an enzyme-linked immunosorbent test was successfully used to reliably identify the plant in infected tomato fruit.

However, this kind of method requires time-consuming, hazardous processes like extraction and derivatization [9]. The results take a long time to get in, and the sample analysis is quite valuable. Future advanced agriculture will undoubtedly require unique research and in-depth use of cutting-edge agricultural technologies.

1.4 Research Question

A. Types of tomato pests in Bangladesh
B. Pest identification
C. Pest analysis
D. Model building using CNN
E. Section of Machine learning algorithms

1.5 Expected Output

Various illnesses and bugs plague tomatoes during their whole growth process. Lack of prompt management will result in a loss in yield or perhaps failure. The most essential thing is to correctly identify illnesses and insect pests. This will help vegetable growers handle diseases and pests more efficiently and increase tomato output. In contrast to the conventional pattern recognition method, the disease and pest recognition method powered by deep learning will input the original image directly. The end-to-end structure is used to change the popular method and address the issue that the feature extractor designed manually struggles to obtain the feature expression that is most similar to the thing's natural attribute. The old technique involved laborious steps like image preprocessing, feature extraction, and classification. With the help of deep learning object identification, it is possible to reduce big losses brought on by illnesses and pests while also delivering a good period of time judgment, which is important for analyzing price and importance. This study may create a dataset of tomato illnesses and pests beneath the significant natural atmosphere utilizing the most current analysis findings of detection theory and deep learning object identification as well as the features of tomato diseases and pests in images.

1.6 Report Layout

This report has six chapters, which are listed below:

A. Introduction

This chapter discusses the study's introduction, objectives, motivation, and expected outcomes, as well as the report's layout.

B. Literature Review

This chapter consists of an introduction, related work, comparative studies, the scope of the problem, and the challenges of the research.

C. Methodology

This chapter begins with a discussion of the research subject and instrument, followed by data collection procedures, statistical analysis, the proposed methodology, and finally the implementation requirements.

D. Experimental results and discussion

This chapter discusses the experimental setup, result analysis, and pest detection.

E. Impact on society, environment, and, sustainability

In this chapter, we have discussed the impact on society, the impact on the environment, ethical aspects, and a sustainability plan.

F. Summary, conclusion, recommendation, and implementation for future research

This chapter consists of a summary of the study, its conclusions, and future work.

CHAPTER 2
Background

2.1 Terminologies

Bangladesh is an agricultural nation where all types of fruits and vegetables may be grown. After potatoes, tomatoes are the second vegetable. Pests cause thousands of tons of tomatoes to be wasted annually. This issue may be resolved and tomato output increased through early pest identification. We have suggested a strategy to pest-proof tomatoes in this study. A dataset of five different pest species—aphids, looper caterpillars, red spider mites, white flies, and thrips-was utilized in this study.

2.2 Related Works

Although crop diseases pose a significant threat to food security, it is difficult to identify them due to a lack of infrastructure. One of the intriguing research topics in the computer and agricultural fields is detective work and identifying sickness from the photos of the plant [4]. Various diseases and pests affect tomatoes at various stages of their growth. Lack of prompt management will lead to a decrease in yield or perhaps failure. It is crucial to find a strategy to control diseases and pests successfully so that vegetable growers may increase tomato yields. Accurately identifying diseases and pests is also of utmost importance [8]. Because of their unique organoleptic, nutritional, and integrative capabilities, tomatoes and their concentrations are essential food items of remarkable culinary and industrial relevance. Numerous types of food adulteration are frequently suspected in various tomato-based products, posing serious economic and frequently even health risks to farmers, the food industry, and consumers [2]. The model achieves a maximum classification accuracy of 0.95 and is built on convolutional layers from the VGG16 and extra layers that are constructed as requested. The accuracy of police work and pest classification findings are vital in desegregating the established models to a reconnaissance mechanism that may accept a photo as an input and recognizes the type of pest in this picture. The use of pesticides can be reduced and agricultural output can be increased through early pest identification. This methodology sought to identify and forecast the typically modest levels of adulterants in ingredient victimization, such as paprika seed, corn starch, sucrose, and salt. During our tests, NIR spectroscopy proved to be more accurate and simpler to use in comparison

[2]. In this study, optical and ultrasonic sensors are employed to track the elimination of food-related fouling, and a neural network regression model is created to forecast how much fouling will still be present on the surface. The outcome suggests that sensors and machine learning techniques may be used together to efficiently monitor improvement processes [3].

The primary goal of this research is to assist farmers with early disease detection and plant prevention. The choice of CNN was made because, in contrast to other classification algorithms, it extracts a large number of possibilities from the picture datasets. An associate degree of accuracy of 97% is attained by the trained model [4]. In order to prevent failures in quantity and for the manufacturing of an agricultural factory-made product, disease classification in plants is essential. Utilizing various Machine Learning and Image processing techniques, the problems in the agricultural industry are identified [6]. A segmentation method that is mostly based on Support Vector Machine is envisaged for this endeavor. Additionally, the severity of any process is determined [6]. This study uses kinetic modeling to examine how tomato paste's color changes when it is heated. The ingredient was created to be processed at three different temperatures—60, 70, and 80 °C—for 25, 50, 75, and 100 minutes, respectively, in order to study the dynamics of color degradation. Total color distinction, saturation index, and hue angle were calculated using three main color parameters—lightness, redness-greenness, and yellowness-blueness values. In this paper, here used the Yolo V3 model to improve the detection accuracy and speed and pests of tomatoes accurately and quickly. The key technology of tomato cuss image recognition in natural surroundings is now available, which provides a reference for intelligent recognition and engineering applications of plant diseases and pest detection [8].

2.3 Comparative Analysis and Summary

Various works have been done in pest detection in many countries. Some of the researchers have done different machine-learning techniques to measure accuracy. One of the popular methods was image recognition using the CNN model. And some other techniques are also introduced such as the YOLO V3 model and so on. Some groups of researchers have done research on leaves also. As far as my knowledge, no study was done using a customized model of CNN and measuring accuracy in Bangladesh like ours.

2.4 Scope of the program

The study is concerned with real and some online collected images that were collected from two small villages of Narayanganj and Kishoreganj. The purpose of the project is to detect the pest at an early stage so that farmers can take the action in early stage to prevent the pest effect and increase the production of tomatoes. This study can be a novel approach to consumers, producers, and intermediaries.

2.5 Challenges

The collection of the datasets presented the major barrier to our investigation. It was challenging to get the ideal picture of a pest. also the time when the data collection was preprocessed. We conducted a number of tests to build the architecture and get the highest accuracy. Many experiments end up being unsuccessful. We finally found a solution. The selection of appropriate algorithms for comparative accuracy was another challenge for us.

CHAPTER 3

Research Methodology

3.1 Research Subject and Instrumentation

Bangladesh is an agricultural country that grows practically every form of fruit and vegetable, including tomatoes. It is a popular and widely consumed vegetable. Tomatoes are commonly grown in Bangladesh as a fruit or vegetable. There are several pests and illnesses that harm tomatoes throughout their whole life cycle, making it critical to detect and diagnose them. Pest infestations in crops cost farmers a lot of money every year, significantly impacting their livelihood. In this study, we built a model that uses a convolutional neural network to identify pests and reduce product loss. In our research, we are dealing with five different types of pests.

Table 3.1: Types of tomato pest

Name	Image
Red spider mite	
Whitefly	
Thrips,	

Looper caterpillars	
Aphids	

3.2 Data Collection Procedure

Data gathering is an essential component of all research activities. Data is a vital component to consider while attempting to learn more about a specific subject, study, research, or even someone. As a result, it is regarded as an essential component of every system that now sustains our world. We must understand the type of information necessary, how to gather it, and where to obtain it. To gather information, both observation and online resources can be employed. Today, data actually serves a diverse set of goals and applications. Regardless of whether you're considering a digital transformation, data collection is a critical stage that should never be overlooked. This is especially true if you want to collect data, make estimates, and manage your business operations in a way that adds significant value. Depending on the sort of activity, data might be gathered quantitatively or qualitatively. Although methods and objectives may differ depending on the field, the primary data-gathering strategies employed are often the same. In other words, careful adherence to particular criteria is necessary to ensure reliable data collection.

As a result, before commencing the data gathering procedure, we reevaluated and reviewed each of our research objectives. To begin, make a list of all of your goals. We've developed some

questions and started gathering data. We acquired information for our study using both online platforms and firsthand observation. 70% of the data was gathered through internet sources, with the remaining 30% obtained from observations. We created a process diagram to show how we should proceed in order to achieve our aim. The workflow diagram is shown below in figure.

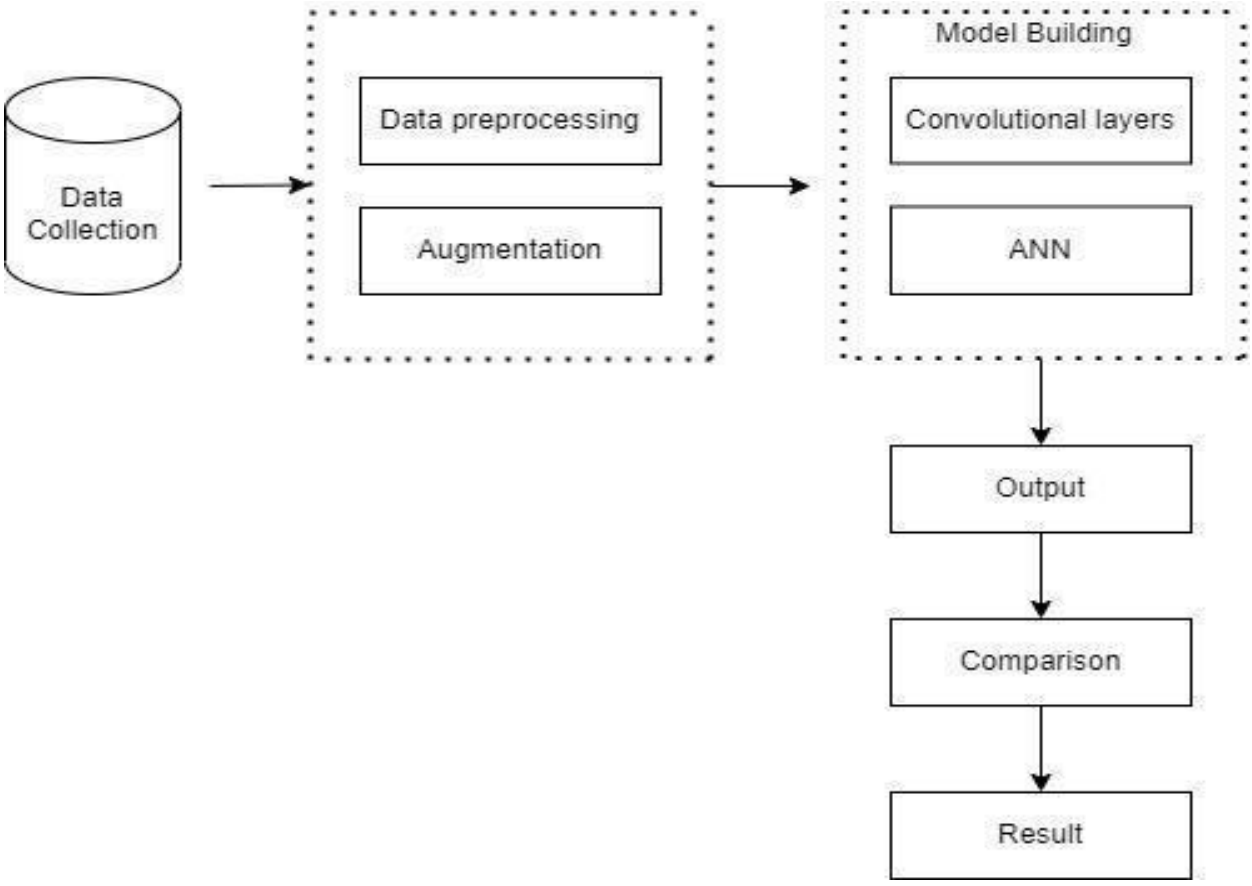


Figure 3.2: Workflow diagram

The graphic above depicts the process of our investigation. First, we gathered data, preprocessed it, and then supplemented it all to meet the demands of the research.

Then, in the following stage, we developed a model, from which we generated output. We then compared them to arrive at the final outcome.

The Sony Alpha A6000, which boasts a 24.3-megapixel APS-C image sensor, a BIONZ X image processing engine, a one-touch remote and sharing function, and a high-resolution OLED Tru-Finder feature, was used for offsite photo collection.

Offline data are collected from two little villages called Gobnagar and Dikdhair in Narayanganj and Kishoreganj. A city in Bangladesh's heart is called Narayanganj. Located 16 kilometers southeast of Dhaka, the nation's capital, it has a population of over 2 million and is part of the Narayanganj District. The metro area of the city is Bangladesh's sixth-largest. The Bangladeshi province of Dhaka includes the Kishoreganj district. There was formerly a Mohkuma in the Mymensingh region. A total of 2495.07 sq km of territory from the Mymensingh district was transferred to create the present-day Kishoreganj District. The district of Kishorganj consists of eight municipalities, thirteen upazilas, 105 union parishads, 39 wards, 145 mahallas, 946 Mouzas, and 1775 villages.

A. Data Preprocessing

One of the main criteria for a research study is the preparation of the data. Preprocessing is putting raw data into a format that computers and machine learning algorithms can understand and evaluate throughout the data mining and analysis phases. Processing data have a major and significant impact on the result. Even the greatest algorithms may not be enough to fully understand the importance of clean, preprocessed data, and machine learning models trained on unneeded data may even make analysis more challenging. They produce an ineffective, completely unhelpful product. Data preprocessing was a challenging endeavor. For each research effort, data cleansing is crucial and a need. The data that was collected during fieldwork proved difficult to tidy up. We removed out-of-focus, superfluous, and irrelevant images from our data gathering. 40% of our data set came from another online source, while 60% of it was obtained by fieldwork. The data must then all be set to the same size in order for the result to be created appropriately. We have five different sorts of pests to work with, thus the cleaned data is organized into folders. Thus, it is ready to implement and generate the output.

B. Image Augmentation

Data augmentation is a technique for artificially increasing the amount of data by generating new data points from the existing ones. The data may be slightly modified or machine learning techniques may be used to add new data points to the generative model of the original data in

order to increase the dataset. A large amount of training data is needed for deep networks in machine learning to operate efficiently. When a machine learning model is trained, its parameters are really optimized so that it can convert a certain input, such as an image, into a specific output. A photograph may be flipped both horizontally and vertically using the flip technique. Additionally, alterations are brought about by rotations of 45, 90, 270, and 180 degrees. The photos were then cropped, blurred, sharpened, and improved for brightness and darkness. There were 11,424 datasets generated. As a consequence, data points are produced for the subsequent steps.

Figure 3.3: Flipping, rotation, and other techniques applied to the image

3.3 Statistical Analysis

Acquiring and analyzing data on five pests—aphids, looper caterpillars, thrips, whiteflies, and red spider mites—in order to identify patterns and trends is known as statistical analysis. It is a technique for using numerical analysis to eliminate bias from data analysis. This method aids in survey and study planning, statistical model construction, and data gathering for research interpretation. Each image in our collection has a width, height, and depth of 40 pixels (RGB channel). The model is summarized in the section below.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 38, 38, 96)        2688
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 19, 19, 96)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 19, 19, 256)       221440
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 9, 9, 256)         0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 9, 9, 384)         885120
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 4, 4, 384)         0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 4, 4, 512)         1769984
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 2, 2, 512)         0
_________________________________________________________________
flatten (Flatten)            (None, 2048)              0
_________________________________________________________________
dense (Dense)                (None, 64)                131136
_________________________________________________________________
dense_1 (Dense)              (None, 5)                 325
=================================================================
Total params: 3,010,693
Trainable params: 3,010,693
Non-trainable params: 0
```

Figure 3.4: Layers of model architecture

3.4 Proposed Methodology

Acquiring and analyzing data on five pests—aphids, looper caterpillars, thrips, whiteflies, and red spider mites—in order to identify patterns and trends is known as statistical analysis. It is a technique for using numerical analysis to eliminate bias from data analysis. This method aids in survey and study planning, statistical model construction, and data gathering for research interpretation. Each image in our collection has a width, height, and depth of 40 pixels (RGB channel). The model is summarized in the section below. The dataset has also been altered by increasing its brightness and blackness. In the second step of the described procedure, pictures were additionally cropped, sharpened, and blurred. CNN is one of the most efficient deep-learning neural network methods in the field of computer vision. CNN is excellent in the fields of image recognition, image classification, picture segmentation, and so forth. Processing several forms of data is CNN's main objective, including 1D processing for signals or time series data, 2D processing for images or audio signals, and 3D processing for movies or depth images.

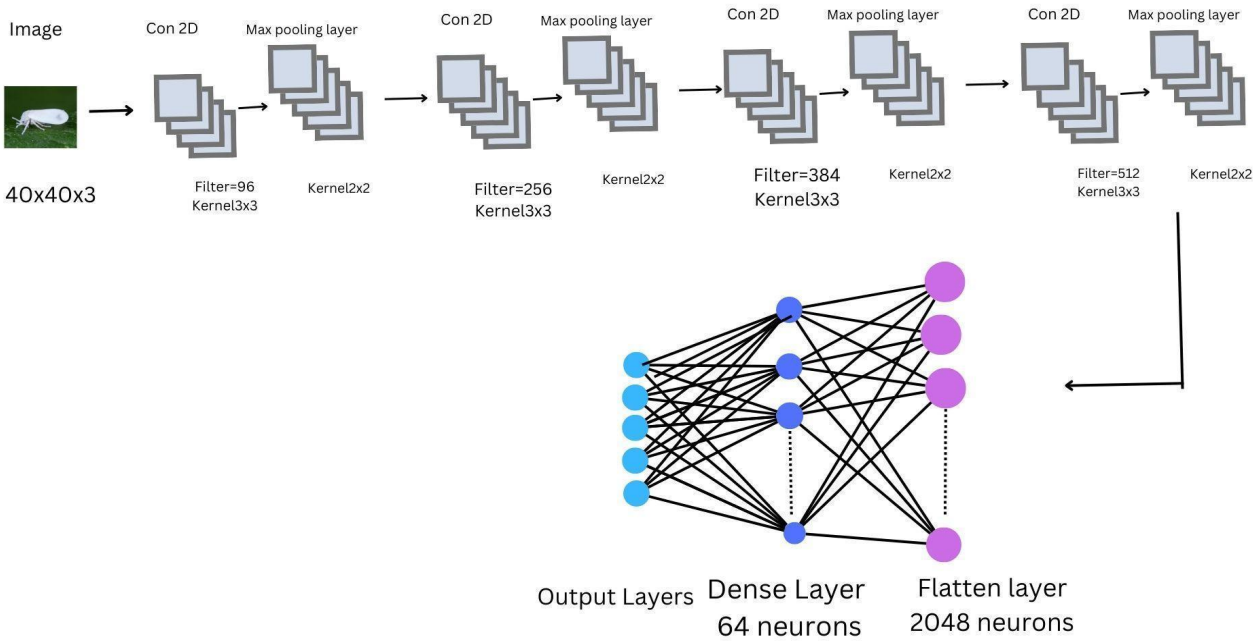


Figure 3.5: Proposed Model Architecture

In our study, we proposed a model that uses a convolutional neural network for feature extraction, which consists of a convolutional layer with 96 filters and kernel size 3x3, followed

by a pooling layer and a convolutional layer with 256 filters and kernel size 3x3 activation functions, and a pooling layer and also 2x2 kernel size of convolution layer. Following that is a second convolution layer with 384 filters, the same padding, a kernel size of 3x3, activation relu, and a pooling layer of 2x2. Another convolution layer follows, with 522 filters, the same padding, a kernel size of 3x3, activation-relu activation, and a pooling layer of 2x2. Following that flatten layer, there are 64 neurons in the dense layer (each neuron contains activation relu), followed by 5 neurons in the output layer with activation softmax. Last but not least, sparse categorical cross-entropy is employed as a loss function and Adam is used as an optimizer with a learning rate of 0.001.

The approach involves training on 80% of the data. Tests are conducted on the remaining 20% of the dataset. Then, we ran 25 epochs to see the outcome graph and produce the result since, otherwise, it appears to be overfitted. As a result, the probability for each class is determined. In the end, we conducted a comparison using three classifiers from Machine Learning. One of the most important properties of the Random Forest Algorithm is its capacity to handle data sets with both continuous variables, as in regression, and categorical variables, as in classification. While Support Vector Machine (SVM) is a supervised machine learning method that may be applied to issues involving classification or regression. For classification problems, it delivers better results. The accuracy of CNN is 94.74 percent, that of RF is 94.09%, that of SVM is 93.34%, and that of K-NN is 94.09%.

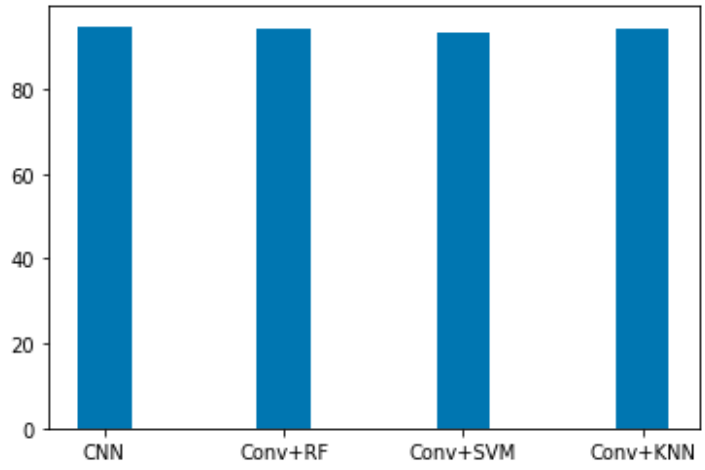


Figure 3.6: Accuracy performance result from the dataset

3.5 Implementation Requirements

The dataset selection is crucial to our investigation. So, those working in the agriculture industry must assist us. To get the photos we need for the research, I also need to train a photographer.

CHAPTER 4
Experimental Result and Discussion

4.1 Experimental Setup

The only way a research study will be successful is if the experiment is correctly conducted. Therefore, achieving a goal needs a good and sophisticated experimental setting. To do this in our study, we have employed picture data. Our dataset was acquired through both online and offline sources. We have utilized a number of Python packages, including TensorFlow, matplotlib, and NumPy. Python is a well-liked computer programming language used to build websites and applications, automate procedures, and do data analysis. Since Python is a general-purpose language, it may be used to create a wide range of applications and isn't specifically designed to address any particular problems. We carried out our experiment on the website Kaggle. Kaggle is a wonderful platform for researchers to experiment with the data, and explore experiments, which helps encourage researchers to do research.

4.2 Experimental Results & Analysis

The result is generated using three algorithms of ML and CNN. With a deep learning foundation, Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) have all been used to identify tomato pests. The recognition model is evaluated using the five assessment metrics. Formulas are given below:

Precision = TP / (TP + FP)

Precision (also known as positive predictive value) is the percentage of relevant examples among the instances that are retrieved, whereas recall is the percentage of relevant instances that were really recovered (also known as sensitivity). In this way, precision and recall are built on the foundation of relevance.

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

Accuracy reveals how frequently the ML model was overall right. The model's precision measures how well it can forecast a certain category. How frequently the model was able to identify a certain category is shown by the recall.

Sensitivity = TP / (TP + FN)

Sensitivity is the capacity of a test to identify a diseased person as positive. A highly sensitive test produces fewer false negative findings and, as a result, misses fewer illness cases.

Specificity = TN / (TP + FP)

This formula repesents the label that is unknown which is negative is known as specificity.

G-mean = (Sensitivity Specificity)

The geometric mean, often known as the G-mean, is the result of class-wise sensitivity. This measure aims to maximize accuracy in each class while preserving a balance between the two. The square root of the sensitivity and specificity in binary classification is the G-mean.

In this context, TP means True Positive, TN means True Negative, FP means False Positive and FN means False Negative is used.

Artificial neural networks of the sort known as CNNs, or convolutional neural networks, are extensively employed for the recognition and categorization of objects and images. Deep Learning can therefore recognize items in a picture by using a CNN. A wide range of activities and functions, including voice recognition in NLP and image processing issues, computer vision tasks including localization and segmentation, video analysis, and identifying obstacles in self-driving cars, all make substantial use of CNNs. Because they are crucial in these quickly evolving and new disciplines, CNNs are well-liked in deep learning[18],[19].

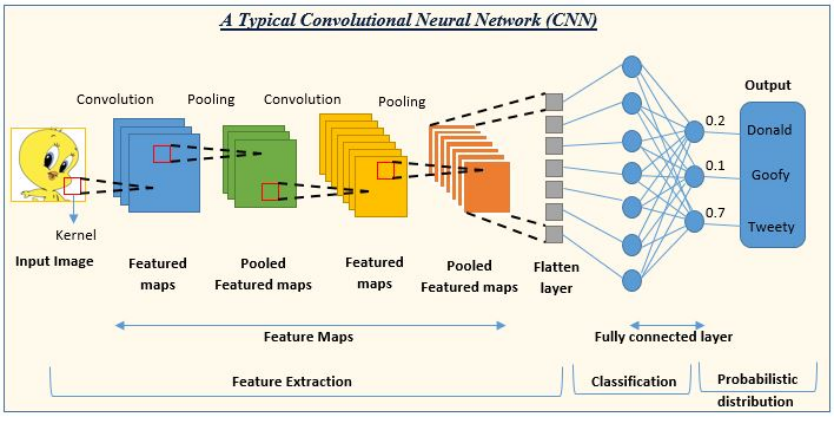


Figure 4.1: A typical convolutive neural network

Descriptions of classifiers are given below-

I. Random Forest classifier: In the supervised learning approach, the well-known machine learning algorithm Random Forest is included. It may be used for ML problems that involve both classification and regression. It is founded on the idea of ensemble learning, a method for combining many classifiers to address challenging issues and enhance model performance.

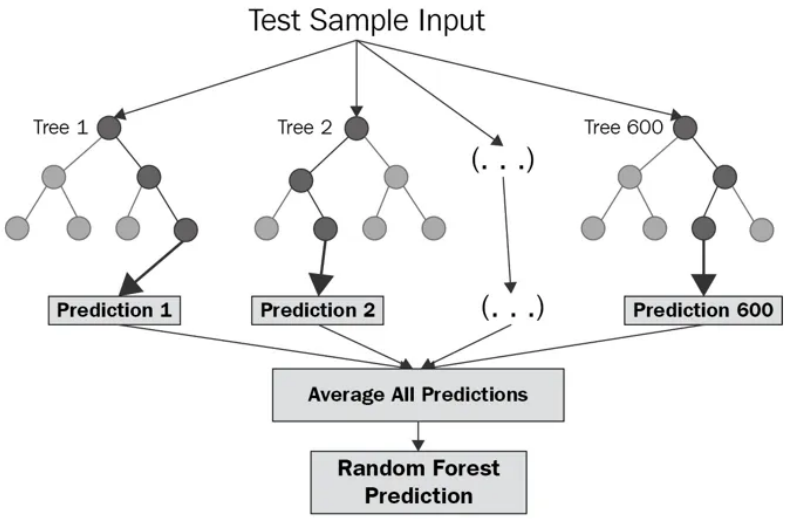


Figure 4.2: Random Forest Classifier

II. K Nearest Neighbour classifier: k-Means KNN is a supervised learning algorithm used for classification, whereas Clustering is an unsupervised learning technique used for clustering.

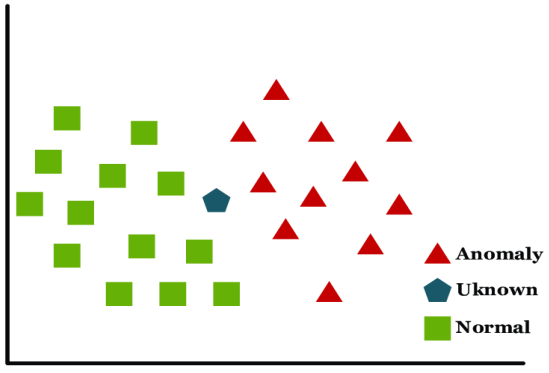


Figure 4.3: K Nearest Neighbour Classifier

III. Support Vector Machine classifier: SVMs are used in a variety of applications, including web pages, intrusion detection, face recognition, email categorization, gene classification, and handwriting recognition. SVMs are used in machine learning for a variety of reasons, including this. On both linear and non-linear data, classification and regression are supported.

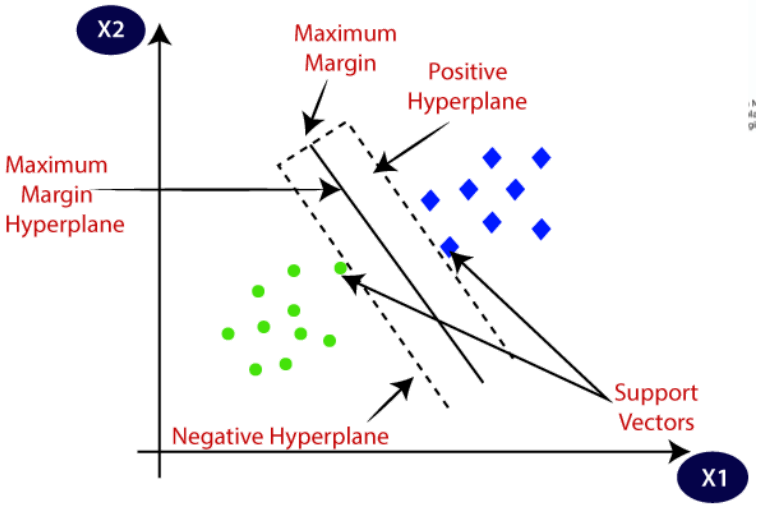


Figure 4.4: Support Vector Machine

Python libraries are run in order to identify related tasks. 80% and 20% of the photos were used to train and test the model, respectively.

Table 4.1: Performance analysis using CNN for five pests

Name of Pests	Precision	Recall	F1	Support
Aphids	0.98	0.87	0.92	461
Looper caterpillars	0.95	0.95	0.95	487
Thrips	0.94	0.95	0.94	493
Whitefly	0.89	0.99	0.93	455
Red spider mite	1.00	0.99	0.99	389

Table 4.2: Performance analysis using Convo+RF for five pests

Name of Pests	Precision	Recall	F1	Support
Aphids	0.93	0.89	0.91	461
Looper caterpillars	0.92	0.92	0.94	487
Thrips	0.93	0.92	0.93	493
Whitefly	0.94	0.96	0.95	455
Red spider mite	1.000	0.97	0.99	389

Table 4.3: Performance analysis using Convo+K-NN for five pests

Name of Pests	Precision	Recall	F1	Support
Aphids	0.92	0.90	0.91	461
Looper caterpillars	0.96	0.93	0.94	487
Thrips	0.89	0.96	0.92	493
Whitefly	0.95	0.96	0.95	455
Red spider mite	1.000	.97	0.99	389

Table 4.4: Performance analysis using Convo+SVM for five pests

Name of Pests	Precision	Recall	F1	Support
Aphids	0.90	0.90	0.90	461
Looper caterpillars	0.95	0.93	0.94	487
Thrips	0.88	0.95	0.91	493
Whitefly	0.95	0.93	0.94	455
Red spider mite	1.000	0.97	0.98	389

The table 4.1 to 4.4 shows the analysis result of the dataset of five pests. In the mentioned tables, precision, recall, f1, and support score are given as the outcome of the implementation.

Table 4.5: Performance result of the CNN and other techniques

	CNN	**Convo+ RF**	**Convo+ SVM**	**Convo+ K-NN**
Accuracy	95.49%	94.09%	93.34%	94.09%
Precision (avg)	95.2%	94.4%	93.6%	94.4%
Specificity (avg)	95.0%	93.8%	92.8%	93.4%
Sensitivity (avg)	95.0%	93.7%	93.4%	94.4%
G-mean (avg)	92.1%	91.6%	92.0%	92.0%

Table 4.5 shows the performance of three different classifiers and also the CNN. The best accuracy is 94.74% for CNN, compared to 894.0%, 93.09%, and 93.34% for convo+RF, convo+KNN, and convo+SVM, respectively. CNN also succeeds in having the greatest sensitivity, specificity, accuracy, and g-mean of 92.1%, 91.6%, 92.0%, and 92.0%.

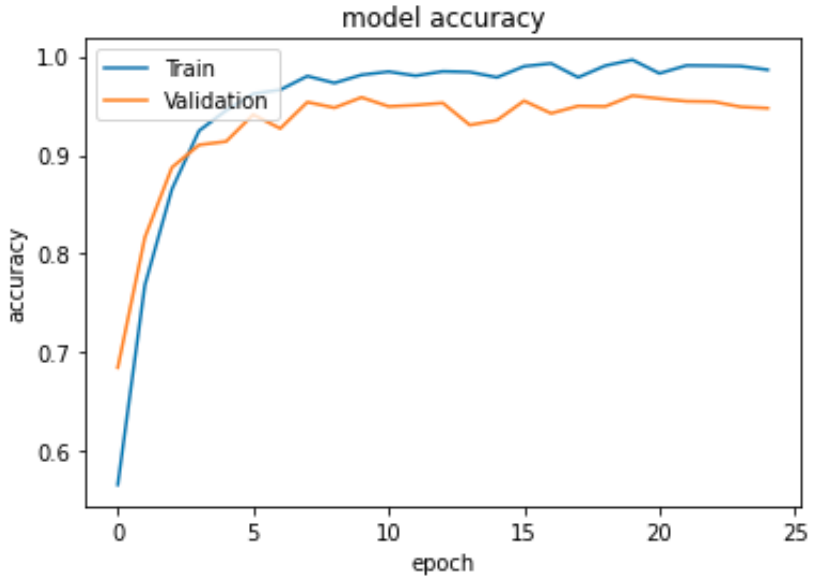


Figure 4.5: Training accuracy vs validation accuracy

From the graph we can see that the maximum validation accuracy of 99.6% was accomplished by CNN during the training accuracy.

4.3 Discussion

The identification of pests in tomatoes is a novel solution. Millions of tomatoes were directly impacted every year, and farmers incurred losses. Our research is thoroughly explained using statistical data and our developed proposed model. Our study's findings were astounding. Our findings were contrasted with those of other techniques. Our research yields excellent outcomes. We believe that our study will have a big impact on society.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

The tomato is one of the most well-known and important crops in the world. In 2018, it produced 188 million tons globally, but a sharp rise in consumption in China and India has led to an even higher level of demand globally. Tomatoes have a very beneficial composition that goes along with their good organoleptic qualities, which is explained by their incredibly high antioxidant content. The carotenoids in tomatoes that are most common are lycopene and lutein, respectively [20]. Tomatoes may be used to make a wide range of products thanks to their distinct technological and practical characteristics.

Regular tomato consumption helps diabetics control their blood sugar and wards against cancer. As well as that, it can support stable blood pressure. Tomatoes contain significant carotenoids like lutein and lycopene. These can protect the eye from light-related damage. To consume more tomatoes, add them to wraps, sandwiches, sauces, or salsas. Instead, consume them boiled or stewed since these methods of preparation might boost the availability of vital nutrients. Tomatoes are among the top ten fruits and vegetables for pesticide residue levels. Wash tomatoes before eating them. Tomatoes are one of the plant meals that have the greatest nutrients. Tomatoes are no exception to the astonishing benefits of eating a variety of fruits and vegetables[23],[24]. As the amount of plant-based foods in the diet increases, the risk of developing heart disease, diabetes, and cancer decreases. There are several methods to prepare tomatoes and they come in a range of sizes and shapes. Among these are cherry tomatoes, stewed tomatoes, raw tomatoes, soups, juices, and purees. Different types may offer various health benefits. For example, cherry tomatoes have more beta-carotene than regular tomatoes. High fruit and vegetable consumption are also linked to better skin and hair, increased energy, and weight loss[25]. Increased fruit and vegetable consumption considerably lowers the risk of obesity and total mortality.

Bangladesh is an agricultural nation where any fruit or vegetable may be grown with ease. Every aspect of a place, including the soil, moisture, and atmosphere, is crucial to the growth of any

sort of fruit or vegetable. Pests cause farmers to lose a lot of money every year. Therefore, if we can find the pest early on, we can boost tomato growth and reduce economic loss. So, we might conclude that our research has a big impact on society.

5.2 Impact on Environment

The ability to effectively manage pests is essential in agricultural operations like farms. Pest species are a significant source of worry due to the possibility of financial loss from crop destruction as well as the serious risk of serious property, machinery, and equipment damage if left untreated. Rodents, birds, and insects are all examples of pest species. Bugs are a severe threat. There are primarily two types of damage they do to crops as they develop. The first of these two forms of damage involves insects eating leaves and drilling holes in stems, fruit, and/or roots[16],[18]. Indirect damage is the second kind of injury that insects can do; in this case, the insects themselves cause little to no harm but instead infect a crop with bacteria, viruses, or fungus. Aphids are a major culprit in this regard since they transmit illness from plant to plant and can spiral out of control as their population rises.

According to various research, landscape simplification reduces the diversity and number of natural enemies in agroecosystems, but its influence on natural pest management is still unknown. Crop pest management by natural enemies is an important ecosystem function that increases crop yield and provides agriculture with a critical but difficult-to-measure ecosystem service. So, controlling pests at an early age can protect the environment in many aspects.

5.3 Ethical Aspects

Early insect identification may aid farmers in increasing tomato yield. It has the potential to boost their economic sector as well as their level of living. Tomatoes are the most common vegetable crop consumed worldwide, behind potatoes and sweet potatoes. They are also an important part of a healthy diet and rank first on the list of canned veggies. In recent years, consumption has increased at a 3% annual pace on average. Tomatoes are now cultivated on 6.1% of all vegetable land, both in the winter and summer. It is cultivated all across the country because of its resilience to a wide range of soil and temperature conditions. So, the people of

Bangladesh will be blessed if pests can be detected in the early stages and proper action is taken to minimize their effects and increase production to satisfy their needs.

5.4 Sustainability Plan

In our research, we proposed a unique method to identify tomato pests early in their life cycles. With static data, we have discussed our phases. Our model has good precision. We believe that our concept may greatly benefit society and will last for a longer time.

CHAPTER 6

Summary, Conclusion, Recommendations, And Implementation For Future Research

6.1 Summary of the Study

Bangladesh has a big agriculture industry, and its economy is increasingly based on agriculture. Plant diseases and pests provide substantial difficulties to the agriculture business. An accurate and timely identification of plant diseases and pests is required in order to develop an early treatment technique and greatly reduce economic losses[16]. We offer a CNN-based technique for identifying pests in tomato plants utilizing both internet datasets and photos captured on-site by camera devices in this study. We have developed a model that shows the best result for the given dataset. CNN is one of the most successful deep-learning neural network approaches in computer vision.

A research methodology is a way for a researcher to outline how they intend to perform their inquiry. CNN's major purpose is to handle different sorts of data, such as 1D signals or time series data, 2D images or audio signals, and 3D films or depth images. A three-layer NN layer comprises a feature extraction convolutional neural network, a pooling layer, and another that utilizes the output as input. Finally, we conducted a comparison using Random Forest Classification. CNN achieves 945.49% accuracy, whereas Random Forest achieves 94.09%, SVM achieves 93.34%, and KNN achieves 94.09%. According to testing results, our proposed system can successfully recognize five unique species of pests (Aphids, Looper caterpillars, Thrips, Whitefly, and Red spider mite), and it can manage difficult conditions that may develop from a plant's surroundings.

6.2 Conclusions

CNN is employed in this work to categorize and detect tomato pests. In the comparison-based study, tomato pest identification is categorized using Random Forest (RF). The series of operations was achieved by training these networks across a number of epochs using a range of learning settings. In our study, our given dataset has shown good performance results. The

development of a CNN-based application can help farmers take precautionary action after harvest.

6.3 Implication for Further Study

Plant pests are a serious problem for the agriculture industry. As a result, early and accurate pest identification and classification may help with prevention efforts while dramatically reducing economic losses. Using advanced technology may enhance the accuracy of prediction. In the future, we may change the layer and compare it with different classifiers to gain better accuracy and reduce financial loss.

REFERENCES

[1] Lotfi, A., Izadi, Z., & Ghasemi-Varnamkhasti, M. (2021). Implementation of a portable electronic tongue system for detection of sodium benzoate adulteration in tomato paste. *Innovative Food Technologies*, *8*(2), 295-308.

[2] Ghajarbeygi, P., Rahimi Niaraki, A., Soltani Abkenar, A., Mahmoudi, R., Jalilevand, F., Sadeghi Niaraki, A., & Alizadeh, A. (2022). Assessment of Sodium Benzoate and Potassium Sorbate Preservatives and Artificial Color in Bulk Tomato Paste Samples in Qazvin, Iran. *Journal of Chemical Health Risks*, *12*(3), 501-507.

[3] JAHAN, S. (2021). *RISK ASSESSMENT OF SELECTED PESTICIDE RESIDUES IN TOMATO AND BITTER GOURD COLLECTED FROM KISHOREGANJ DISTRICT OF BANGLADESH* (Doctoral dissertation, DEPARTMENT OF AGRICULTURAL CHEMISTRY, SHER-E-BANGLA AGRICULTURAL UNIVERSITY, DHAKA-1207).

[4] Ahmed, S., Rahman, T., Ripon, M. S. H., Rashid, H. U., Kashem, T., Md Ali, M. S., ... & Daud, Z. A. M. (2021). A food frequency questionnaire for hemodialysis patients in Bangladesh (BDHD-FFQ): development and validation. *Nutrients*, *13*(12), 4521.

[5] Ahmed, S., Rahman, T., Ripon, M. S. H., Rashid, H. U., Kashem, T., Md Ali, M. S., ... & Daud, Z. A. M. (2021). A food frequency questionnaire for hemodialysis patients in Bangladesh (BDHD-FFQ): development and validation. *Nutrients*, *13*(12), 4521.

[6] Ahmed, S., Rahman, T., Ripon, M. S. H., Rashid, H. U., Kashem, T., Md Ali, M. S., ... & Daud, Z. A. M. (2021). A food frequency questionnaire for hemodialysis patients in Bangladesh (BDHD-FFQ): development and validation. *Nutrients*, *13*(12), 4521.

[7] Dutta, N. K., Sen, R., Faruk, M. I., & Islam, M. A. (2021). Impact of Agricultural Chemical Inputs on Human Health and Environment in Bangladesh. *Agricultural Chemical Inputs*, 22.

[8] Yasmin, N. S., Ali, M., Robbani, M., & Rajib, M. M. R. (2022). Forecasting Tomato Variety for Different Seasons and Regions of Bangladesh. *American Journal of Multidisciplinary Research and Innovation*, *1*(1), 1-8.

[9] Shalaby, A. A., Seloma, A. S. O., & Shalaby, M. A. (2022). Study of Bifenazate, Indoxacarb and Emamectin Benzoate Residues on Tomato. *Journal of Plant Protection and Pathology*, *13*(7), 169-174.

[10] Kotsanopoulos, K. V., & Uddin, M. N. (2022). Authenticity of Tomato. In *Authenticity of Foods of Plant Origin* (pp. 83-104). CRC Press.

[11] Yegrem, L., & Dagnaw, L. A. (2022). Pretreatments, dehydration methods and packaging materials: effects on the nutritional quality of tomato powder: a review.

[11] Fusco, G. M., Burato, A., Pentangelo, A., Cardarelli, M., Nicastro, R., Carillo, P., & Parisi, M. (2022). Microbial Consortium Applications Can Affect Quality and Primary Metabolism of Processing Tomato.

[12] Comlekcioglu, N., Simsek, M., Hayoglu, I., & Kiroglu Zorlugenc, F. (2018, August). Responses of processing tomato (Solanum lycopersicum Mill) to regulated deficit irrigation under semi-arid conditions: yield and quality. In *XXX International Horticultural Congress IHC2018: International Symposium on Water and Nutrient Relations and Management of 1253* (pp. 213-220).

[13] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science*, *156*(3), 312-322.

[14] Meshram, V., Patil, K., Meshram, V., Hanchate, D., & Ramkteke, S. D. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences*, *1*, 100010.

[15] Panigrahi, K. P., Sahoo, A. K., & Das, H. (2020, June). A cnn approach for corn leaves disease detection to support digital agricultural system. In *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)* (pp. 678-683). IEEE.

[16] Pocock, M. J., Marzano, M., Bullas-Appleton, E., Dyke, A., De Groot, M., Shuttleworth, C. M., & White, R. (2020). Ethical dilemmas when using citizen science for early detection of invasive tree pests and diseases. *Management of Biological Invasions*, *11*(4), 720-732.

[17] Vatti, S.R., 2020. *Detecting Pests on Tomato Plants using Convolutional Neural Networks* (Doctoral dissertation, Dublin, National College of Ireland).

[18] Vitalis, F., Zaukuu, J.L.Z., Bodor, Z., Aouadi, B., Hitka, G., Kaszab, T., Zsom-Muha, V., Gillay, Z. and Kovacs, Z., 2020. Detection and quantification of tomato paste adulteration using conventional and rapid analytical methods. *Sensors*, *20*(21), p.6059.

[19] Simeone, A., Woolley, E., Escrig, J. and Watson, N.J., 2020. Intelligent industrial cleaning: A multi-sensor approach utilising machine learning-based regression. *Sensors*, *20*(13), p.3642.

[20] Vimal Kumar, M.N., Shobana Nageswari, C., Saran Kumar, E., Sadhiya, S. and Thiyagarajan, J., 2022. Early-Stage Plant Disease Prediction Using Multilayer Convolutional Neural Network Framework. *JOURNAL OF ALGEBRAIC STATISTICS*, *13*(3), pp.233-240.

[21] Velioğlu, H.M., Boyacı, İ.H. and Kurultay, Ş., 2011. Determination of visual quality of tomato paste using computerized inspection system and artificial neural networks. *Computers and electronics in agriculture*, *77*(2), pp.147-154.

[22] Kadam, K.U., 2020. Identification of groundnut bud necrosis virus on tomato fruits using machine learning based segmentation algorithm. *Int J Fut Gener Commun Netw*, *13*, pp.259-264.

[23] Ganje, M., Jafari, S.M., Farzaneh, V. and Malekjani, N., 2018. Kinetics modelling of color deterioration during thermal processing of tomato paste with the use of response surface methodology. *Heat and Mass Transfer*, *54*(12), pp.3663-3671.

[24] Liu, J. and Wang, X., 2020. Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Frontiers in plant science*, *11*, p.898.

Final Report

ORIGINALITY REPORT

20%	13%	7%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	www.researchgate.net Internet Source	2%
4	Jun Liu, Xuewei Wang. "Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network", Frontiers in Plant Science, 2020 Publication	1%
5	www.mdpi.com Internet Source	1%
6	link.springer.com Internet Source	1%
7	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	1%

8	Submitted to Technological Institute of the Philippines Student Paper	1%
9	lycopene.com Internet Source	1%
10	Submitted to University of Witwatersrand Student Paper	1%
11	Submitted to Deer Valley Unified School District Student Paper	1%
12	Submitted to University of Wales Institute, Cardiff Student Paper	1%
13	Mohammad Ganje, Seid Mahdi Jafari, Vahid Farzaneh, Narges Malekjani. "Kinetics modelling of color deterioration during thermal processing of tomato paste with the use of response surface methodology", Heat and Mass Transfer, 2018 Publication	<1%
14	Submitted to Shelton State Community College Student Paper	<1%
15	en.wikipedia.org Internet Source	<1%
16	Submitted to University of Strathclyde Student Paper	

		<1%
17	www.v7labs.com Internet Source	<1%
18	www.medicalnewstoday.com Internet Source	<1%
19	Submitted to City University College of Science and Technology Student Paper	<1%
20	Submitted to University of Surrey Student Paper	<1%
21	mibcom.ase.ro Internet Source	<1%
22	Submitted to University of Hertfordshire Student Paper	<1%
23	citeseerx.ist.psu.edu Internet Source	<1%
24	Submitted to University of Mpumalanga Student Paper	<1%
25	Nikolaos Sideris, Georgios Bardis, Athanasios Voulodimos, Georgios Miaoulis, Djamchid Ghazanfarpour. "Using Random Forests on Real-World City Data for Urban Planning in a Visual Semantic Decision Support System", Sensors, 2019 Publication	<1%

26	beei.org Internet Source	<1%
27	Submitted to The College of New Jersey Student Paper	<1%
28	documentation.sas.com Internet Source	<1%
29	John, Vijay, Seiichi Mita, Zheng Liu, and Bin Qi. "Pedestrian detection in thermal images using adaptive fuzzy C-means clustering and convolutional neural networks", 2015 14th IAPR International Conference on Machine Vision Applications (MVA), 2015. Publication	<1%
30	www.biorxiv.org Internet Source	<1%
31	Juanping Zhao, Weiwei Guo, Bin Liu, Zenghui Zhang, Wenxian Yu, Shiyong Cui. "Preliminary exploration of SAR image land cover classification with noisy labels", 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017 Publication	<1%
32	Submitted to University of North Texas Student Paper	<1%
33	Srie Raam Mohan, Syed Saqib Bukhari, Andreas Dengel. "Layout Error Correction	<1%

Using Deep Neural Networks", 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), 2018

Publication

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		