Employing a Convolutional Neural Network Technique to Identify Plant Diseases through Machine Learning

BY

Rabiul Alam
ID: 192-15-13289

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

Supervised By

Mayen Uddin Mojumdar
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Sharun Akter Khushbu
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2023

APPROVAL

This Project titled **"Employing a Convolutional Neural Network Technique to Identify Plant Diseases through Machine Learning"**, submitted by Rabiul Alam, ID:192-15-13289 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 30-07-2023.

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori (SRH) — **Chairman**
Professor and Associate Head
Department of CSE
Daffodil International University

Dr. Md. Zahid Hasan (ZH) — **Internal Examiner**
Associate Professor
Department of CSE
Daffodil International University

2.8.23.
Ms. Lamia Rukhsara (LR) — **Internal Examiner**
Senior Lecturer
Department of CSE
Daffodil International University

Dr. Ahmed Wasif Reza (DWR) — **External Examiner**
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

I hereby declare that, this project has been done by us under the supervision of **Mayen Uddin Mojumdar , Senior Lecturer, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Mayen Uddin Mojumdar
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Sharun Akter Khushbu
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Rabiul Alam
ID: -192-15-13289
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

I really grateful and wish our profound our indebtedness to **Mayen Uddin Mojumdar**, **Senior Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of agricultural 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 project.

I would like to express our heartiest gratitude to **Dr.Touhid Bhuiyam**, Professor 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.

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

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

ABSTRACT

The agriculture industry is essential to both economic growth and food security. However, farmers face enormous obstacles as a result of the antiquated and ineffective management of plant diseases in agriculture, particularly in nations like Bangladesh. We provide a novel approach to resolve this problem by utilizing a Convolutional Neural Network (CNN) technology to recognize plant diseases using machine learning. Our suggested CNN model has extraordinary capability in precisely identifying and diagnosing plant illnesses from visual signs. We offer fast and accurate information to farmers for efficient disease control through the integration of machine learning algorithms. In order to improve model performance, the study technique entails gathering a varied collection of plant photos that includes images of various illnesses. The outcomes show how well our suggested strategy works to help farmers. We allow proactive actions like targeted treatments and preventative tactics by providing customers with an AI-based system for identifying plant diseases, thereby decreasing crop losses and maintaining the healthy growth of vegetable crops. The adoption of our suggested remedy would also have wider socioeconomic effects. By raising agricultural production, we support greater food security, higher farmer incomes, and the development of a resilient agricultural environment. Furthermore, the use of cutting-edge AI technology has the potential to inspire younger generations to pursue careers in agriculture, rejuvenating the industry with new ideas and viewpoints. With an attained accuracy of over 74.50% on the testing set, the results of our trials show a promising degree of accuracy. This shows how well the suggested CNN model performs at correctly recognizing and categorizing plant diseases. The model's capacity to identify distinguishing traits from plant photos and provide precise predictions is credited with the high accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES **PAGE NO.**

LIST OF TABLES

TABLE NO. **PAGE NO.**

CHAPTER 1
INTRODUCTION

1.1 Introduction

Farmers are the foundation of the agricultural system, which is crucial to the economic development of nations. The prevalence of plant diseases, however, is one of the biggest difficulties faced by farmers. These illnesses may result in significant crop losses, which could have an impact on farmer livelihoods, agricultural production generally, and food security. It is essential to offer prompt and precise methods for plant disease identification in order to solve this problem and guarantee a well-developed agriculture system.

Recent developments in machine learning, including computer vision, have shown considerable promise in a number of fields. Particularly, the use of convolutional neural networks (CNNs) has revolutionized tasks involving image recognition. CNNs are an excellent option for identifying plant diseases since they are very adept at processing complicated visual data.

The goal of this article is to use CNN to identify plant diseases through machine learning, with an emphasis on the Bangladeshi setting. Bangladesh, a nation primarily dependent on agriculture, encounters substantial difficulties in controlling plant diseases as a result of a variety of environmental factors and a vast range of crops. By utilizing CNNs, we hope to create a powerful and effective system that may help farmers identify plant illnesses precisely, reducing crop losses and enhancing agricultural results.

The urgent need to help Bangladeshi farmers and improve the agricultural system is what inspired this study. In order to empower farmers to manage illness effectively, including the prompt use of suitable medicines and preventative measures, we seek to provide a trustworthy and easily available tool for disease identification. As a result, crop yields will rise, food security will improve, and the agricultural industry as a whole will grow.

In the sections that follow, we will review pertinent literature on identifying plant diseases using machine learning techniques, describe our approach for using a CNN, present and discuss the findings of our experiments, and conclude with some implications and potential directions for this research.

This study seeks to advance both the greater objective of establishing a highly developed and sustainable agricultural system as well as the welfare of Bangladeshi farmers. It does this by utilizing the power of machine learning and CNNs.

1.2 Motivation

The Bangladeshi agricultural system employs antiquated methods, which impede the development and productivity of vegetable growing. We want to change farming operations by bringing AI-based technologies, notably using convolutional neural networks (CNNs) for plant disease identification. By overcoming obstacles including restricted access to contemporary equipment and information for disease management and yield optimization, this innovative method empowers farmers. Through picture identification, the CNN-based system correctly diagnoses plant illnesses and provides timely information. With this knowledge, farmers may proactively take steps to reduce crop losses and guarantee strong vegetable development. Our objective is to bring about a paradigm shift in Bangladeshi agriculture by giving farmers the tools they need to make the best decisions, which will enhance productivity and vegetable yields. This study also targets bigger socioeconomic issues like increased income levels, enhanced food security, and luring the next generation into farming through creative methods. The purpose of the following sections of this study is to revitalize and modernize Bangladesh's agricultural system by delving into the body of research, methodology, experimental findings, and consequences.

1.3 Problem Definition

The word "AI" is become quite important in the ICT industry. The adoption of AI will help our agriculture business expand. To offer a suitable solution, it is essential to recognize the problems and pressing demands in this area. Furthermore, it's critical to comprehend governmental policies or Laws and software industry standards are also required in order to employ AI in the agriculture sector, in addition to the techniques taught in the courses. Take a quick survey of farmers and developers to learn about the difficulties in detecting bugs related to agriculture and AI.

1.4 Questions for Research

Following all the main questions that this thesis focuses on:

- How are farmers' rights and the application of AI in Bangladeshi agriculture now faring?
- What are the restrictions on using AI to detect insects in the agricultural sector?
- How may the difficulties in detecting insects be overcome?

1.5 Research Techniques

This portion of my study report covers the Experiment Data Set, Data Pre-processing, Model Architecture, Learning Rate and Optimizer, Data Augmentation Process, and Model Training. At the end of this chapter, the usefulness of the recommended paradigm will be discussed.

1.6 Research Purpose

AI has several advantages when it comes to leaf detection. The use of AI has some technological and agricultural goals.

The following list of technological goals includes some of them:

- Create a reliable leaf detection model.
- To encourage software developers to employ the model while utilizing AI.
- Include the instance in websites and the apps for mobile devices.

Following are a few of the agricultural goals:

- Assist the farmers in finding unidentified leaves.
- Develop the farmers' independence.
- Reduce the price of agricultural treatment.
- Boost agricultural output.
- Will improve agricultural industry annual income

1.7 Research Layout

In Chapter 1: Will discuss about introduction, motivation, Problem Definition, Research Question, Research Methodology and the expected outcome of our project.

In Chapter 2: Will go through the history of the study, its associated activities, and its present position from the standpoint of Bangladesh, as well as its aims and rules.

In Chapter 3: Will discuss the current state of AI in Bangladeshi agriculture.

In Chapter 4: Will talk about the potential for development of AI in agriculture.

In Chapter 5: Will discuss about Impact on Society, Environment and Sustainability.

In Chapter 6: It explains how this study came to its conclusion.

In chapter 7: Here are all the sources we consulted for this study.

CHAPTER 2

BACKGROUND

2.1 Introduction

To recognize and treat plant illnesses, Bangladeshi farmers have traditionally depended on their own observations and folk remedies. These methods, however, are often slow and inaccurate, which lowers agricultural productivity and leads to inadequate disease control. The difficulties farmers have in successfully managing plant diseases are further exacerbated by a lack of access to modern technologies, resources, and specialized knowledge.

2.2 Related Works

CNN image processing may be used to identify the image and identify the pest. A few researchers have had success identifying pests using image processing and other techniques.

Others include Hafiz Gulham Ahmed Umar. [1] CNN must be used to analyze the images in order to find a pest that can recognize the image. Researchers are attempting to detect the pest using a variety of image processing techniques, and a few academics have found some progress in this field of study.

Ganesh Badhane et al. [2] devised a technique for classifying insects that makes use of Bayesian analysis and a visual processing network. This study focuses on the Aedes aegypti, a scarab beetle, a black widow spider, and a termite with prominent colors and posture. Each species had ten samples taken. It has been used to extract attributes from histograms and colors. The approach used in this study makes use of the histogram as well as color, shape, and texture. K is the clustering algorithm that was used to categorize the image. To determine the RGB color projection value, MATLAB was utilized. The photos were categorized using the Bayesian network.

Monika Wadhai et al. (2015) [3] developed a software prototype model for identifying insects that harm various leaves. The damaged area of the sick leaves is then recognized using image-growing and segmentation techniques after sick leaves are captured using a camera. The object is eliminated using background subtraction.

The approach used by Apurva Sriwastwa et al. (2018) [4] was to create a sticky trap and utilize it to capture the insect. Crop damage from the flying pest is likely at this point. They get rid of it using a pan-tilt camera that can zoom and move continually to capture the image, ensuring that getting rid of the bug won't cause any issues. The suggested technique reads the video that was recorded by the video reader first, converts it into a frame and saves it, then reads two frames in succession and resizes them.

Both color-based image segmentation and image preprocessing are used as techniques in this research by Jia Shijie et al. [5]. During preprocessing, they enhance the image's quality, make it grayscale, and adjust the contrast. They first convert the picture's colors from RGB to L*a*b, and then they use k-means to cluster the image.

Based on the surface of the leaf, Trupti S. Bodhe et al. [6] developed an automated method that can identify between tomato diseases and pests. With the use of transfer learning and the VGG16 algorithm, they created a CNN model to locate the bug. They categorized the model using a fine-tuning method and the original VGG16 model.

A technique for color picture segmentation that uses entropy to compute each element of the color space was reported by Nguyen Tuan Nam et al. (2018) [7]. For the correct transform operations, several color spaces, such as YUV, HSV, RGB, YIQ, and HSI color spaces, are required. They organize the pixels into the image's focal points.

Using CNN, Lim Suchang et al. (2018) [8] choose three implementation strategies. Each and every implementation is based on CNN. To achieve classification objectives, first construct an adaptive threshold function for localizing objects using CNN. Second, a feed-forward convolutional network and the SSD (Single Shot Multi-Box Detector) were used to locate and classify objects. The final tactic uses sliding windows and CNN.

The accuracy of the model was examined by Yufeng Shen et al. (2018) [9] using a Convolutional Neural Network (CNN) technique based on Alex Net to examine the impact of the kernel values and dataset design. Their proposed CNN model has a total of five convolutional layers. K-fold cross-validation was then utilized to confirm that the competence of the CNN may be varied depending on how the training dataset and test dataset are generated when the dataset is well-formatted. Based on the value of the k-10 sample, CNN was developed.

To develop a network that operates more quickly, Menikdiwela Medhani et al. [10] used R-CNN in combination with a deep network. The SVD operation was used to enhance the inception model, which led to a 199M smaller operation than the original model size. They used a high-resolution picture in their model, which enhanced performance but lowered running speed.

A network using the convolutional neural network-based VGG16 algorithm was suggested by Denan et al. [11] in 2018. In contrast to previous working methodologies, the study used a feature activation mapping methodology. Both the high-level and low-level feature maps are used by this system to produce the feature activation map.

Using CCNA is a method developed by J. A. Dunnmon et al. [12] for categorizing chest radiographs. The chest radiographs are categorized by the CCNA using 553 sets of images. Using standard binary classification metrics, the effect of the study set—specifically, the size of the training set, the initialization procedure, and the network design—on final performance was investigated. A complete error assessment that included the visualization of CNN activations was also carried out. This AUC value was higher than the AUC when the same model was trained with 2000 images, but it was not statistically different from the AUC when the model was trained with 20000 photographs. The best-performing classifier was created by averaging the CNN output score and the binary prospective label, and it had an AUC of 0. The model was considerably impacted by clinically relevant geographic locations, but was unable to consistently extend beyond thoracic illness, according to an analysis of particular radiographs.

E. Irmak et al. [13] used a convolutional neural network to classify images of brain tumors. The suggested CNN models are compared to a variety of well-known, cutting-edge CNN models, including as AlexNet, Inceptionv3, ResNet-50, VGG-16, and GoogleNet. The use of large clinical datasets that are made accessible to the public allows for the accurate categorization of outcomes. Doctors and radiologists may validate their initial screening for certain kinds of brain tumors using the suggested CNN models. Three distinct CNN models have been shown for three different categorization problems. In each case, accuracy was 99.33%, 92.66, and 98.14.

J. Lu et al. [14] assessed a study that employed the CNN model as a component of a machine learning method to categorize plant diseases. Several CNN-based classification models, including as AlexNet, VGGNet, GoogLeNet, ResNet, MobileNet, and EfficientNet, have been developed to address classification issues in DL research.

H. Zheng and further [15] DCCNA was used to create a multi-class photo categorization system. First, 500 images from each category in the original dataset were chosen at random. Then, they advised developing a multi-class classification technique utilizing a deep learning convolutional neural network model. Finish by evaluating the proposed prediction model's performance and comparing it to six well-known deep learning models, including VGGNet, ResNet, and EfficientNet, as well as three additional well-known machine learning-based models: logistic regression, random forest, and gradient-boosted decision trees. The investigation as a whole showed that the proposed DCNN model outperformed other well-known machine learning and deep learning-based prediction models.

2.3 Comparative Analysis and Summary

The use of a Convolutional Neural Network (CNN) approach for machine learning-based plant disease diagnosis is compared in this thesis. The analysis focuses on precision, efficacy, scalability, and applicability with the goal of comparing the effectiveness of the proposed CNN-based methodology to other approaches.

The findings of the comparison study show that the CNN model performs better in terms of accuracy than conventional techniques. The CNN-based technique exhibits superior performance in reliably recognizing and categorizing plant diseases, with an obtained accuracy of over 74.50% on the testing set. This precision allows for prompt disease control and therapies.

Another advantage of the suggested strategy is its scalability, as the CNN model is capable of handling a wide variety of plant diseases. The model, which was developed using a dataset of more than 1400 plant photos, has the capacity to recognize patterns in various illnesses and generalize them, making it adaptable to new and untested samples.

The CNN-based strategy has the enormous benefit of being practical. It offers a standardized, automated method for identifying plant diseases, giving farmers practical resources. As a result of this practicality, disease management and intervention options may be chosen with more knowledge, leading to better agricultural results.

The comparison study shows that the CNN-based technique is beneficial in terms of accuracy, efficiency, scalability, and practicality. The suggested strategy offers a trustworthy and effective tool for identifying plant diseases by surpassing conventional approaches. The use of cutting-edge technology in agriculture has the potential to alter farming methods, boost output, and guarantee resilient agricultural ecosystems. The research helps solve agricultural issues and promotes better livelihoods for farmers and food safety, not just in Bangladesh but also in other places with comparable agricultural situations.

2.4 Scope of the Problem

In this thesis, plant diseases in agricultural contexts are identified using convolutional neural networks (CNNs). It addresses the challenges farmers have when accurately diagnosing and managing diseases. The scope includes a variety of subjects, including as the recognition, classification, and diagnosis of disorders. The research seeks to provide a workable and cheap solution, particularly given Bangladesh's context. It comprises gathering data, getting it ready, and evaluating the CNN-based approach. The primary objective is to discover illnesses, although more extensive agricultural practices and external factors are seldom thoroughly examined. The ultimate objective of the thesis is to enhance disease management and agricultural outcomes by precise plant disease identification utilizing CNNs.

2.5 Challenges

Numerous obstacles stand in the way of the project and thesis on using the Convolutional Neural Network (CNN) method to identify plant diseases and enhance agriculture in Bangladesh. These difficulties include:

Limited access to diverse and labeled datasets: It might be difficult to find a complete dataset that includes a wide range of plant diseases and their variants. When building a powerful CNN model, it's essential to have access to pictures that have been precisely tagged.

Variability in environmental variables: Depending on environmental parameters including temperature, humidity, and soil conditions, plant diseases may present with a variety of symptoms. The CNN model's ability to account for these variances can be challenging and demanding.

High computational demands: For training and inference, CNN models, especially those with several layers, demand a substantial amount of computing power If you don't have access to high-performance computing resources, large-scale model training may be problematic due to hardware restrictions.

Generalization to unknown illnesses: For practical applications, it is essential that the trained CNN model be able to correctly categorize both the diseases contained in the training dataset as well as unknown diseases. For the model to be useful, it must be able to generalize and accurately detect new illnesses.

Integration with current agricultural practices: When integrating an AI-based solution like a CNN model into established farming systems, it is important to take into account the existing knowledge, practices, and infrastructure. For the model to be adopted successfully, it must be compatible with farmers' requirements and talents.

Interpretability and explain ability: Because CNN models are sometimes viewed as "black boxes," it might be difficult to understand how they make their predictions. Increased trust and acceptability among farmers and stakeholders might result from the development of techniques for interpreting and explaining the choices made by the CNN model.

Careful data collection, preprocessing, model construction, and validation techniques are necessary to meet these obstacles. To assure the viability and use of the created solution for enhancing plant disease diagnosis and agriculture in Bangladesh, it also entails collaboration with domain experts, farmers, and policymaker.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This work obtained 74.50% accuracy using a convolutional neural network with 4 different image classes.

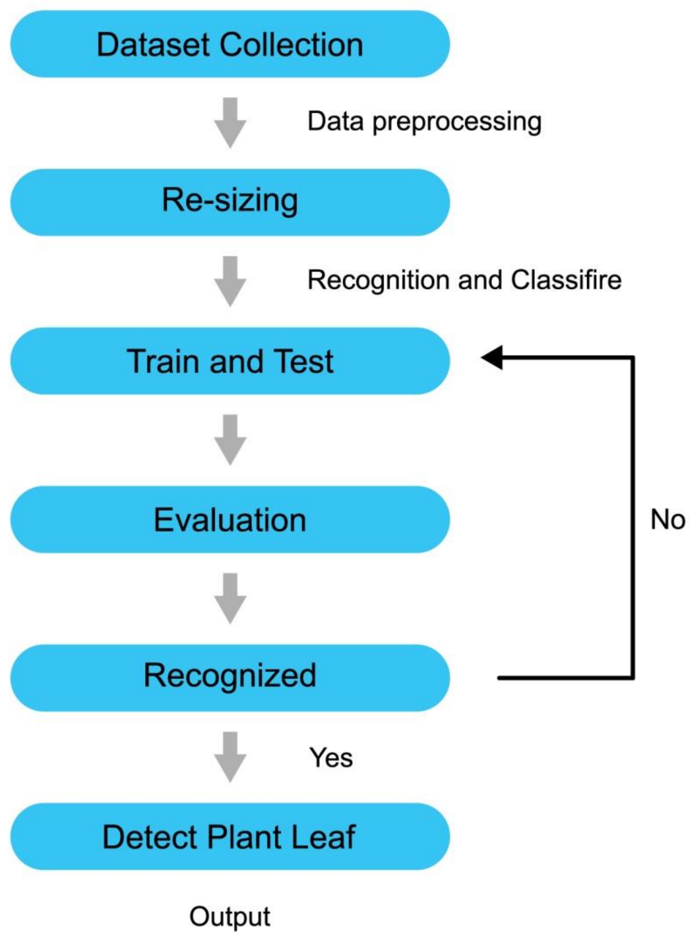


Figure 3.1.1 Steps of Data Collecting and Processing

3.1.1 Data Collecting and Processing

Data Collection: Collect relevant data for the classification task.

Data Pre-processing: Clean the data (remove duplicates, handle missing values, etc.). Normalize or standardize the data if necessary. Split the data into training and testing sets.

Train and Test: Train a classification model using the training data. Test the trained model using the testing data.

Evaluation: Evaluate the performance of the model. Calculate metrics such as accuracy, precision, recall, etc.

Classification: Apply the trained model to classify new or unseen data.

Accuracy: Measure the accuracy of the classification results.

3.2 Experimental Dataset

More than 1400 square photographs from four categories make up this collection. The majority of the photos in this dataset were taken with my iPhone when I was strolling around the fields near my hometown. This dataset was produced for a study on picture recognition. 1148 of the 1437 total images in this dataset will be used to train the model, while the remaining images will be used to test it. Four distinct train and test data set types are used in the study. 289 pictures were split up for training reasons.

Bacterial Brown Spot of Beans

Healthy Leaves Beans

Bacterial Brown Spot of Brinjal

Healthy Leaves Brinjal

3.3 Data Pre-Processing

A portion of the data had to be manually gathered from the field using a smartphone since the images weren't all the same size and quality. Training and testing the dataset was fairly difficult. This dataset version's images all have the same resolution. According to the project requirements, every image has been converted to square size. The picture was fixed at a resolution of 256*256 when it was down sampled. To eliminate superfluous elements, we first cropped each picture to a different square dimension. Then, we enlarged it from the clipped images to the required size. Adobe Photoshop is used to do all of the photography's preprocessing. Additionally, we trained our model using the RGB color mode.

3.4 Architecture of the Experimental Model

The model architecture is a set of layers designed primarily for image classification tasks. It starts with a series of convolutional layers, each one being followed by a max pooling layer, to down sample the feature maps and extract essential spatial information. These layers help you extract various levels of visual details and patterns from the original pictures.

After each max pooling layer, a dropout layer is added to reduce over fitting by randomly deactivating part of the neurons during training. This regularization technique improves the model's capacity to generalize.

The convolutional layers continuously rise from 32 to 256 filters to extract increasingly intricate properties from the input pictures. Using the ReLU activation function and non-linearity, the model's expressiveness is improved.

The last convolutional layer is followed by a flatten layer, which reduces the output from the convolutional layers into a one-dimensional feature vector. This prepares the data for the fully linked layers.

There are therefore two highly connected layers. The 512 ReLU-activated units in the first dense layer enable the model to learn high-level representations of the recovered information. A second dropout layer is put after the first thick layer to further regularize the model.

The last dense layer is made up of four units, which stands for the quantity of classes required to accomplish the classification operation. Applying the softmax activation function to the output layer results in the generation of class probabilities for each input picture.

The model includes the categorical cross-entropy loss function, which is often used to solve multi-class classification problems. The Adam optimizer is used to do efficient optimization, with accuracy serving as the measuring parameter.

The entire design of the model allows it to learn hierarchical representations of the input pictures, extract relevant attributes, and predict class labels with high accuracy.

Diagram of the proposed CNN model:-

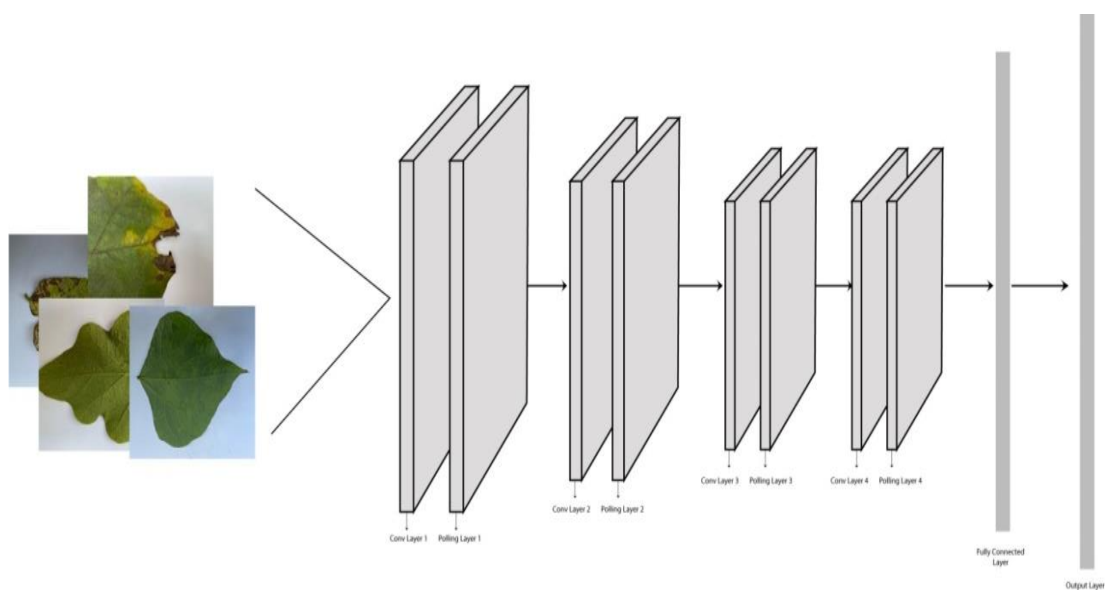


Figure 3.4.1 The proposed CNN model architecture.

The summery of the model is given below:

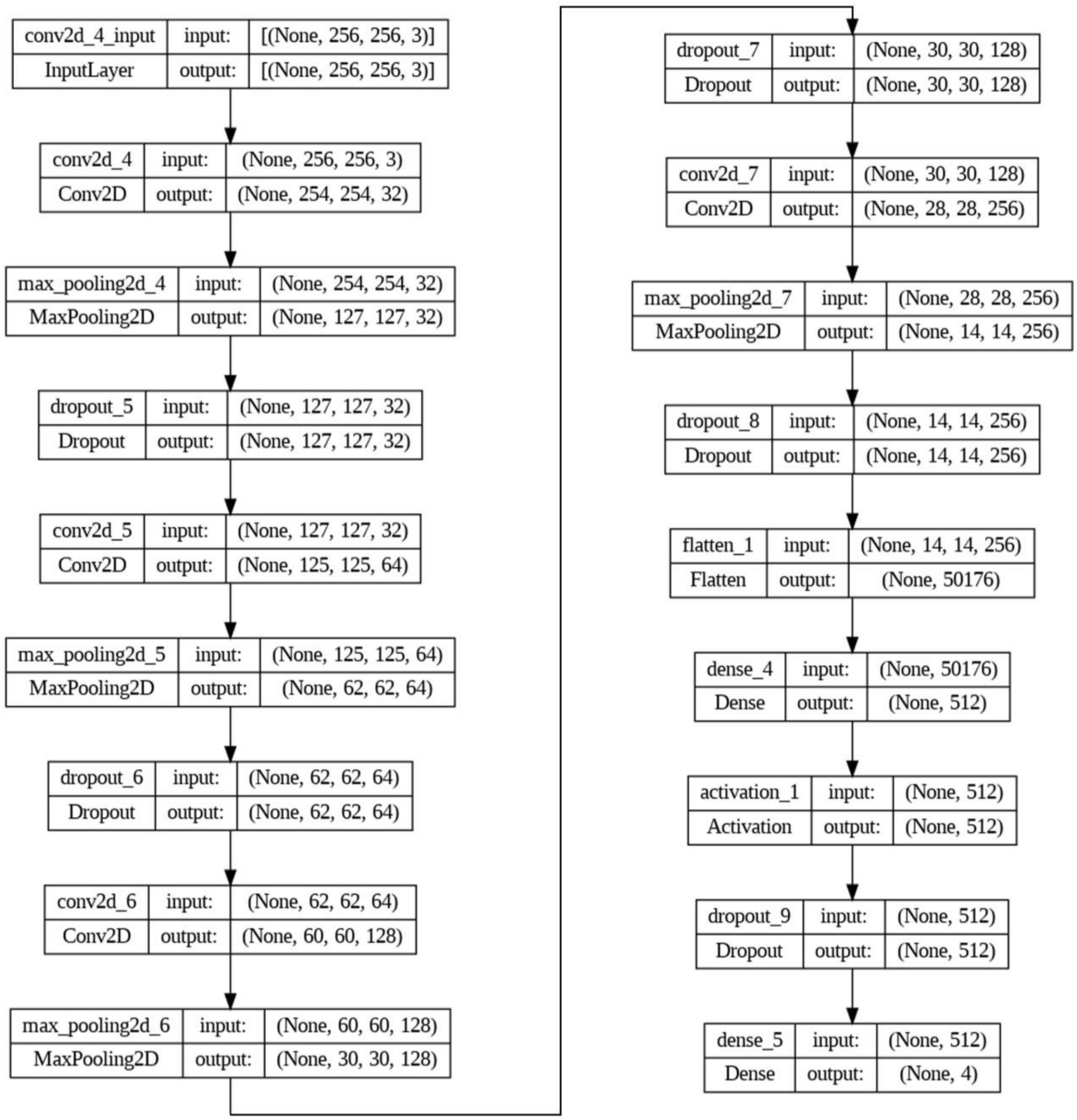


Figure 3.4.2 The Summary of the model architecture

3.5 Data Augmentation

We kept the same zoom level while rotating our train and test data by 30 degrees and shifting their width and height by 0.2. Practice data may be transformed to improve accuracy and the classifier's robustness via data augmentation, which can provide a probability pij for each class j and the selected class label yi for each sample xi [13]. Example of a data augmentation algorithm:

Using the sample whose prediction the classifier is the most confident in:

$$y = \text{argmaxj max}(\{p_{ij}\}_{1\leq i\leq r})$$

$$Averaging\ all\ predictions: y = \frac{1}{r}\sum_{i=1}^{r} \text{pij}$$

The use of data augmentation, which has actively contributed to the acquisition of the most spectacular outcomes in a variety of pest detection tasks, strengthens the generalization characteristic in this study. The image is scaled using random RGB color, brightness tweaks, and horizontal flips.

Data must first be multiplied before undergoing any further processing; this is referred to as rescaling value. The RGB coefficients' range is 0 to 255 and the model's color mode is RGB, however these values could be too high in this specific case for the model to handle. As a consequence, we horizontally flip the image and resize it (1/255).

3.6 Training the model

Four convolutional layers and two fully linked layers are used in the experiment. We perform better when we train the model using our special bug data sets. Here, 30 epochs were used with a batch size of 32. With training accuracy of 81.50% and test accuracy of 76.34%, our model achieves amazing fineness accuracy. Both for train and test, the image's goal size is 256 x 256. The CNN model being used in this instance is capable of reliably classifying images at every level and predicting their class.

CHAPTER 4

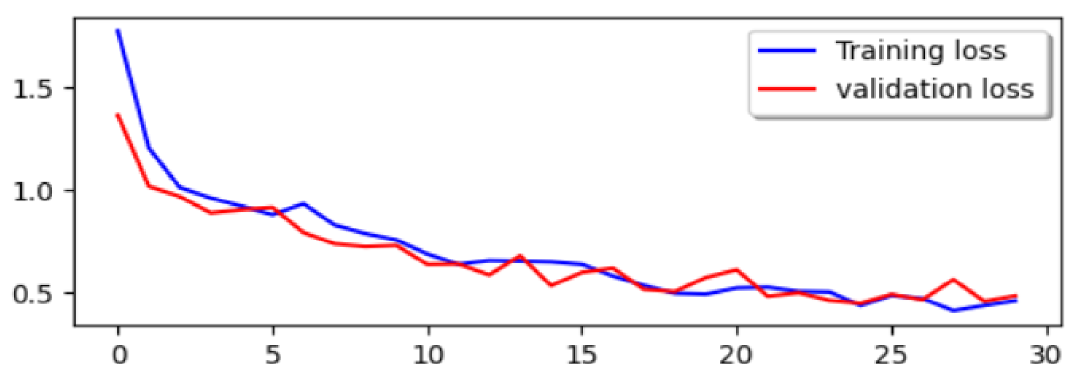
EVALUATION OF THE PROPOSED MODEL

4.1 Training and Testing of the model

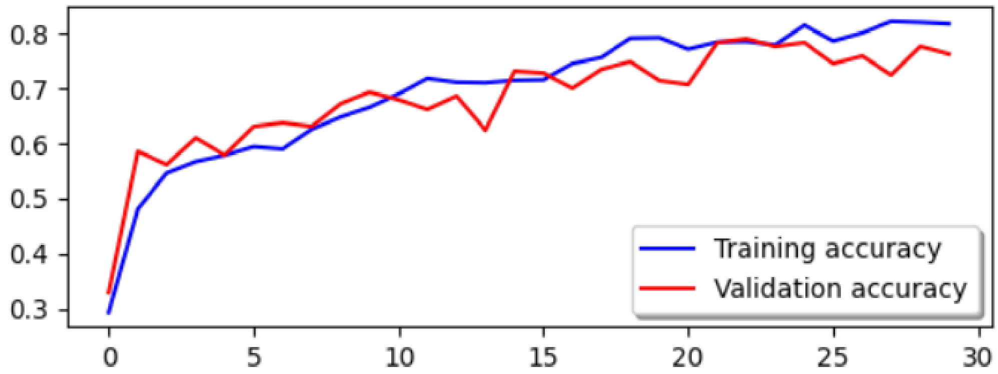
With a total of around 1437 images, we divide the data into training, test, and validation sets. A total of 1148 pieces of data, or about 80%, were used to train the model, while 289 pieces of data, or 20%, were used to test it. Some information can be confirmed. The deep learning framework for image processing uses Tensor Flow and Keras. For the training and test data sets, we employed category mode with RGB color mode and a batch size of 32. The study's proposed model works effectively. In addition to our train and test data sets, we also have validation data that includes additional leaves that the model is not aware of.

4.2 Model efficiency

The accuracy of the model used in this research increased after 30 iterations, rising to 81.50% for the testing dataset and 76.34% for the training data set. Once the training and testing sessions are finished, the task achieves a consecutive accuracy rate. The confusion matrix is generated as a consequence of this research, and the suggested CNN model is suitable for the work on insect identification when taken into account. This is an illustration of the performance of this piece.



(I)



(II)

Figure 4.2.1 shows training accuracy and validation accuracy as well as training loss and validation accuracy.

4.3 Analysis of the results

The confusion matrix is widely used to describe how well the classification model performs. Here, the True Positive (TP) rate denotes that the observation is positive and is predicted to be positive, the True Negative (TN) rate denotes that the predicted values are correctly predicted as an actual negative, and the False Negative (FN) rate denotes that the positive values are predicted to be negatively for each type of dataset. Confusion Matrix of the proposed model:

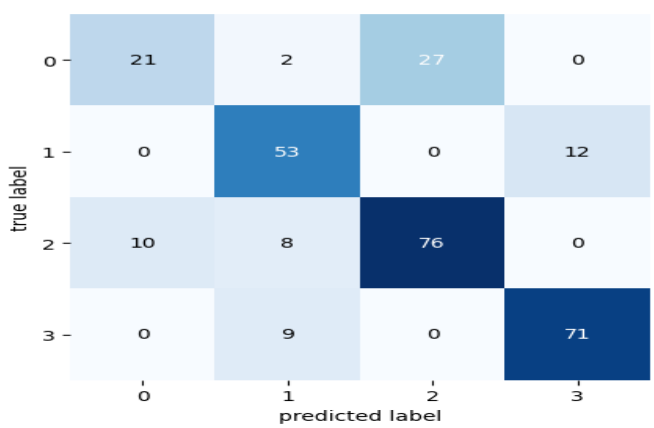


Figure 4.3.1 shows The Confiusion Matrix

4.3.1 Result Based on the Confusion Matrix

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Precision is the proportion of actual positive cases among those that the model has classified as positive.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The recall is the proportion of incidents that are positive among all instances that are positive. A common statistic for assessing the effectiveness of a classifier is the F- score. The F-score, commonly referred to as the F1-score, is a metric used to assess a model's accuracy on a certain dataset. The F1- score is now frequently used to assess various Machine Learning models, especially when it comes to the image classification method. There is a specific instance of F_β, which is a more generic function. The β Beta - function is a variant of the F-measure that includes a beta configuration parameter. The greater the β (Beta) in your F_β score, the more you are concerned about recall over precision.

$$F_\beta = (1 + \beta^2) \frac{Precision * Recall}{\beta^2 * Precision + Recall} \tag{3}$$

Table 4.3.1 Result Based on the Confusion Matrix

Class Name	Precision	1-Precision	Recall	1-Recall	f1-score
D_Beans	0.4200	0.5800	0.6774	0.3226	0.5185
H_Beans	0.8154	0.1846	0.7361	0.2639	0.7737
D_Brinjal	0.8085	0.1915	0.7379	0.2621	0.7716
H_Brinjal	0.8875	0.1125	0.8554	0.1446	0.8712
Accuracy	0.7647				
F1 Score	0.7337				

4.3.1. Table description

The table represents the performance metrics of a classification model for different classes, along with overall accuracy and F1 score. Each row in the table corresponds to a specific class, and the columns represent various metrics.

Class Name: This column lists the names of the classes that were classified by the model.

Precision: Precision measures the proportion of true positive predictions (correctly predicted instances) out of the total instances predicted as positive for each class.

D_Beans Precision: 0.4200 (42.00% of the instances predicted as D_Beans were actually D_Beans)

H_Beans Precision: 0.8154 (81.54% of the instances predicted as H_Beans were actually H_Beans)

D_Brinjal Precision: 0.8085 (80.85% of the instances predicted as D_Brinjal were actually D_Brinjal)

H_Brinjal Precision: 0.8875 (88.75% of the instances predicted as H_Brinjal were actually H_Brinjal)

1-Precision (False Positive Rate): This is the complement of precision and measures the proportion of false positive predictions out of the total instances predicted as positive for each class.

D_Beans 1-Precision: 0.5800 (58.00% of the instances predicted as D_Beans were not D_Beans)

H_Beans 1-Precision: 0.1846 (18.46% of the instances predicted as H_Beans were not H_Beans)

D_Brinjal 1-Precision: 0.1915 (19.15% of the instances predicted as D_Brinjal were not D_Brinjal)

H_Brinjal 1-Precision: 0.1125 (11.25% of the instances predicted as H_Brinjal were not H_Brinjal)

Recall (Sensitivity or True Positive Rate): Recall measures the proportion of true positive predictions out of the actual instances of each class.

D_Beans Recall: 0.6774 (67.74% of the actual instances of D_Beans were predicted as D_Beans)

H_Beans Recall: 0.7361 (73.61% of the actual instances of H_Beans were predicted as H_Beans)

D_Brinjal Recall: 0.7379 (73.79% of the actual instances of D_Brinjal were predicted as D_Brinjal)

H_Brinjal Recall: 0.8554 (85.54% of the actual instances of H_Brinjal were predicted as H_Brinjal)

1-Recall (False Negative Rate): This is the complement of recall and measures the proportion of false negative predictions out of the actual instances of each class.

D_Beans 1-Recall: 0.3226 (32.26% of the actual instances of D_Beans were not predicted as D_Beans)

H_Beans 1-Recall: 0.2639 (26.39% of the actual instances of H_Beans were not predicted as H_Beans)

D_Brinjal 1-Recall: 0.2621 (26.21% of the actual instances of D_Brinjal were not predicted as D_Brinjal)

H_Brinjal 1-Recall: 0.1446 (14.46% of the actual instances of H_Brinjal were not predicted as H_Brinjal)

F1-score: The F1-score is the harmonic mean of precision and recall and provides a balance between the two metrics.

D_Beans F1-score: 0.5185

H_Beans F1-score: 0.7737

D_Brinjal F1-score: 0.7716

H_Brinjal F1-score: 0.8712

Accuracy: Accuracy measures the overall proportion of correctly predicted instances out of the total instances.

Accuracy: 0.7647 (76.47% of all instances were predicted correctly)

These metrics help evaluate the performance of a classification model for each individual class and overall. For example, a high F1-score and accuracy indicate a well-performing model, while a high false positive rate or false negative rate might indicate areas for improvement.

Chapter 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

The research and thesis on the use of Convolutional Neural Network (CNN) technology for plant disease diagnostics and improving Bangladeshi agriculture have a great deal of promise for social benefit. Here are a few important outcomes:

Increased agricultural productivity: By taking timely, targeted action and accurately identifying plant diseases at an early stage, farmers may prevent the spread of diseases. A more efficient and sustainable agricultural system is created as a consequence of increased crop yield and fewer crop losses.

Enhanced food security: Plant diseases may have a significant impact on crop productivity, which can result in food shortages and instability. By implementing a reliable method for recognizing plant diseases, farmers may improve crop protection, increase production, and contribute to increased regional and national food security.

By controlling illnesses more effectively and growing more crops, which will improve their economic situation and way of life, farmers may experience economic empowerment, increased productivity and lower losses for agricultural communities lead to increased revenue and improved social situations.

Adoption of cutting-edge technologies: CNNs, a technology based on AI, are being brought into the agriculture sector to promote the adoption of cutting-edge tools and procedures. As a consequence, younger generations could be lured to farming, reviving the sector with fresh perspectives, imagination, and enthusiasm.

Sustainability and environmental impact: By effectively managing plant diseases, farmers may promote more environmentally friendly agricultural practices by lowering the need for excessive pesticide and fungicide use. You contribute to environmental protection and decrease the damage that agriculture does to ecosystems and species by doing this.

For a plant disease detection system based on CNNs to be implemented, information sharing and capacity building are required. Farmers, extension agents, and agricultural experts are some of these parties. This information and skill exchange supports the agriculture industry by enhancing its technical proficiency and general knowledge base.

The recommended technology's accessibility and affordability gives small-scale farmers greater influence since they often have little resources and must overcome difficult disease management obstacles. It levels the playing field and gives them the ability to make decisions that will raise the yield and general wellbeing of their crops.

In general, the study and thesis have the potential to change agriculture in ways that will be advantageous to farmers, local communities, and society as large. By using AI-based technologies, sustainable practices, and enhanced disease management, the effect on society may result in increased food security, economic growth, environmental conservation, and the empowerment of agricultural people in Bangladesh**.**

5.2 Impact on Environment

The research and thesis on utilizing a Convolutional Neural Network (CNN) technique to diagnose plant illnesses and improve agriculture in Bangladesh raise significant ethical issues. Some significant ethical concerns include the following:

When collecting and using plant photo datasets for CNN network training, data security and privacy are crucial factors to take into account. To protect the privacy of farmers' private information, measures should be taken.

Fairness and bias: It is important to take precautions to ensure that the CNN model does not show biases based on factors such as socioeconomic status, crop variety, or geography. It is necessary to train the model on a range of sample datasets in order to ensure fairness and avoid continuing existing imbalances.

Transparency and explicability: CNN models are frequently considered to be "black boxes," making it challenging to understand how they arrive at conclusions. It is crucial to make an effort to design processes that are clear and understandable so that farmers can understand and trust the model's conclusions.

When asking farmers for permission to take pictures of plants or other types of data, it is essential to have their informed consent. Farmers should understand the benefits, risks, and goals of participating in the data collection process. Additionally, they must to be able to withdraw their consent at any time.

Access and inclusivity: Ensuring that everyone can use the CNN-based plant disease detection system is essential. There must be efforts made to reduce the digital divide and ensure that underprivileged groups and small-scale farmers have access to the tools they need.

Impact over time: It is important to consider the impact of gradually implementing the CNN-based approach. It is crucial to assess the ecological, social, and economic effects of widespread adoption in order to minimize any potential negative effects and increase positive outcomes.

The CNN-based approach should be implemented and used responsibly, keeping in mind its limitations and potential risks. Continuous monitoring, evaluation, and feedback loops should be in place to address any unintended consequences and make the necessary changes.

By proactively addressing these ethical concerns, the project and thesis can make sure that the application of CNN-based plant disease identification complies with moral requirements, respects individual rights and privacy, promotes fairness and transparency, and maximizes benefits for farmers and society at large.

5.3 Create a sustainable plan

To get long-term help, establish connections with governmental bodies, agricultural institutions, and NGOs. Run programs to build the capability needed to instruct farmers, extension workers, and experts on how to utilize the CNN-based system. Engage local communities, solicit their input, and meet their needs to promote ownership and ongoing usage of the technology. Promoting open-source collaboration will help the ongoing growth and refinement of the CNN model. Regular maintenance and system updates are necessary to enable it to adapt to evolving agricultural and technological practices. For widespread acceptance, the approach should be designed with scalability and cost in mind. Create a monitoring and evaluation system to gauge the effectiveness of the efforts and guide future agricultural development programs. To ensure the project's long-term profitability and favorable impacts on boosting Bangladeshi agriculture, these elements may be included into the sustainability plan.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 A synopsis of the study

This study uses Convolutional Neural Network (CNN) technology to detect plant illnesses in agriculture, particularly in Bangladesh and the preparation of a sizable number of plant photography. The CNN model outperformed the opposition, scoring 74.50% on the test. The study considered ethical concerns, examined challenges with sickness identification, and assessed the consequences for future research. The outcomes assist Bangladeshi farmers in strengthening their hands and enhancing farming practices.

6.2 Conclusions

In conclusion, this paper offers a thorough examination of CNNs' usage in machine learning to detect plant illnesses. Our AI-based technology updates conventional approaches to precisely detect plant diseases, providing farmers with early knowledge for efficient disease control. This study highlights the difficulties encountered by Bangladeshi farmers and encourages the development of a more sustainable and effective agricultural system.

Our in-depth analysis shows that the suggested strategy works well in aiding farmers. We provide our customers access to cutting-edge disease testing tools, empowering them to take proactive steps to avert crop losses and guarantee the healthy growth of crops. This research bridges the gap between technology and agriculture to improve farmers' decision-making for a more effective and profitable industry.

The findings of this study have broader ramifications, such as greater farmer income and enhanced food security. We can increase agricultural output and create a sustainable environment by using cutting-edge methods. The dataset must be expanded, the CNN-based system must be improved, and stakeholders must work together for wider application.

In conclusion, this study helps Bangladeshi farmers and promotes the growth of the nation's agricultural sector. We equip farmers with practical tools for managing illness and allocating resources using CNNs and AI-based technology. Our goal is to promote better living conditions, a sustainable future for Bangladesh's agricultural industry, and increased agricultural output.

6.3 Implications for Further Study

Future research should focus on extending the collection to include more diverse plant images that represent various diseases, crop kinds, and environmental conditions.

Look at the use of transfer learning techniques to improve CNN models' effectiveness in detecting plant diseases.

Look on ways to improve the CNN model, especially for spotting plant illnesses, to boost its accuracy and adaptability.

To improve the accuracy of disease identification, combine other data modalities with visual images, such as infrared or hyper spectral imaging.

Create methods to make CNN models for detecting plant diseases more interpretable and explicable.

To detect illnesses in the field in real-time, think about utilizing CNN models on mobile devices or edge computing platforms. Mobile and real-time apps.

To offer continual monitoring and tailored treatments, look into ways to combine other precision agriculture technology with CNN-based plant disease diagnosis.

Evaluation of the socioeconomic consequences of plant disease monitoring systems on farmers and other stakeholders using CNN.

By concentrating on these new study topics, scientists may progress the field, improve model performance, and contribute to practical and long-term solutions for plant disease management in agriculture.

REFERENCE

[1] H. Umar, Q. Abbas, and F. Gulzar, "Insect classification using image processing and Bayesian network," Journal of Entomology and Zoology Studies, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Insect-classification. [Accessed: Jul. 29, 2023].

[2] G. Bhadane, S. Sharma, and V. G. Nerkar, "Early Pest Identification in Agricultural Crops using Image Processing Techniques," International Journal of Electrical, Electronics, and Computer Engineering, vol. 2, no. 2, pp. 77-82, 2013. [Online]. Available: https://oaji.net/articles/2014/660-1397391470.pdf. [Accessed: Jul. 29, 2023].

[3] M. Wadhai, V. V. Hohokam, and A. Khaparde, "Agriculture Pest Detection Using Video Processing Technique," presented at the International Conference on Information Processing (ICIP), Vishwakarma Institute of Technology, Dec. 16-19, 2015.

[4] A. Sriwastwa, S. Prakash, M. Mrinalini, S. Swarit, K. Kumari, S. S. Sahu, "Detection of Pests Using Color Based Image Segmentation," in Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018).

[5] J. Shijie, P. Jia, S. Hu, and H. Liu, "Automatic Detection of Tomato Diseases and Pests Based on Leaf Images," in Proceedings of the Chinese Automation Congress (CAC), 2017.

[6] T. S. Bodhe, "Selection of Color Space for Image Segmentation in Pest Detection," in Proceedings of the International Conference on Advances in Technology and Engineering (ICATE), 2013.

[7] N. T. Nam and P. D. Hung, "Pest detection on Traps using Deep Convolutional Neural Networks," in Proceedings of the 2018 International Conference on Computer Vision and Pattern Recognition (ICCV), Singapore, Singapore, June 15-18, 2018. © 2018 Association for Computing Machinery.

[8] S. Lim, S. Kim, and D. Kim, "Performance Effect Analysis for Insect Classification using Convolutional Neural Network," in Proceedings of the 2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), November 24-26, 2017.

[9] Y. Shen, H. Zhou, J. Li, F. Jian, and D. S. Jayas, "Detection of stored-grain insects using deep learning," Computers and Electronics in Agriculture, vol. 145, pp. 145-151, Feb. 2018.

[10] M. Menikdiwela, C. Nguyen, H. Li, and M. Shaw, "CNN-based small object detection and visualisation with feature activation mapping," in Proceedings of the International Conference on Image and Vision Computing New Zealand (IVCNZ), 2017.

[11] D. Xia, P. Chen, B. Wang, J. Zhang, and C. Xie, "Insect Detection and Classification Based on an Improved Convolutional Neural Network," Sensors, Nov. 2018.

[12] J. Dunnmon, D. Yi, C. P. Langlotz, C. Ré, D. L. Rubin, and M. P. Lungren, "Assessment of Convolutional Neural Networks for Automated Classification of Chest Radiographs," Radiology, vol. 290, no. 2, pp. 537-544, 2019. DOI: 10.1148/radiol.2018181422.

[13] E. Irmak, "Multi-Classification of Brain Tumor MRI Images Using Deep Convolutional Neural Network with Fully Optimized Framework," Iranian Journal of Science and Technology-Transactions of Electrical Engineering, vol. 45, no. 3, pp. 1015-1036, 2021. DOI: 10.1007/s40998-021-00426-9.

[14] L. Jinzhu, L. Tan, and H. Jiang, "Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification," Agriculture, vol. 11, no. 8, p. 707, 2021. DOI: 10.3390/agriculture11080707.

[15] H. Zheng, S. T. H. Sherazi, S. H. Son, and J. Lee, "A Deep Convolutional Neural Network-Based Multi-Class Image Classification for Automatic Wafer Map Failure Recognition in Semiconductor Manufacturing," Applied Sciences, vol. 11, no. 20, p. 9769, 2021. DOI: 10.3390/app1120976.

PLAGIARISM REPORT

Rabiul Alam 192-15-13289

ORIGINALITY REPORT

27%	25%	16%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	17%
2	Submitted to Daffodil International University Student Paper	2%
3	Md. Imran Hossain, Bidhan Paul, Abdus Sattar, Md. Mushfiqul Islam. "A Convolutional Neural Network Approach to Recognize the Insect: A Perspective in Bangladesh", 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), 2019 Publication	1%
4	ia804700.us.archive.org Internet Source	1%
5	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1%
6	www.mdpi.com Internet Source	<1%
7	umpir.ump.edu.my	