

CNN-Based Disease Detection for Cauliflower

By

FARIHA AHMED

ID: 191-15-12304

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised by

Shah Md. Tanvir Siddiquee

Assistant Professor

Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Mayen Uddin Mojumdar

Senior Lecturer

Department of Computer Science and Engineering
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

Dhaka, Bangladesh

January 13, 2025

APPROVAL

This Project titled “CNN-based Disease Detection for Cauliflower”, submitted by Name: Fariha Ahmed, ID No :191-15-12304, 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 13 January, 2025.

BOARD OF EXAMINERS



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

Chairman



Most. Hasna Hena (HH)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Md. Ferdouse Ahmed Foysal (FAF)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



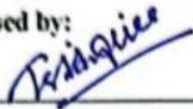
Dr. Md. Arshad Ali (DAA)
Professor
Department of Computer Science and Engineering
Jagannath University

External Examiner

DECLARATION

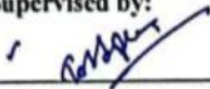
We hereby declare that this project has been done by us under the supervision of **Shah Md. Tanvir Siddiquee, Assistant Professor, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:



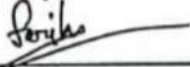
Shah Md. Tanvir Siddiquee
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:



Mayen Uddin Mojumdar
Senior Lecturer
Department of Computer Science and Engineering
Daffodil International University

Submitted by:



Fariha Ahmed
ID: 191-15-12304
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. I am deeply grateful to everyone who has assisted me in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **s Shah Md. Tanvir Siddiquee, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Deep learning specifically a subfield called machine learning** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express our heartfelt gratitude to Professor Dr. Touhid Bhuiyan, the Head of the Department of Computer Science and Engineering, for his kind help in finishing my project and to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

I would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

Convolutional Neural Networks (CNN) have received a lot of attention for disease identification in agricultural crops like cauliflower because of its capacity to evaluate visual patterns in images automatically and reliably. This paper investigates a CNN-based framework for identifying and classifying diseases that affect cauliflower leaves, using deep learning techniques to process and evaluate images of infected and healthy plants. By training the CNN on a dataset of annotated photos, the model is able to extract significant properties such as texture, color, and shape abnormalities caused by diseases such as fungal infections, bacterial spots, or vitamin shortages. The suggested system detects illnesses with high accuracy, providing a fast, reliable, and scalable option for farmers and agronomists. This technology has the potential to improve crop management, reduce losses, and promote sustainable agriculture practices by allowing for early diagnosis and action. The suggested method is intended to help farmers and agricultural professionals diagnose diseases more quickly, reliably, and affordably. Early detection leads to timely response, lowering crop losses and pesticide consumption. This promotes sustainable agricultural practices. This study demonstrates the potential of CNNs to change disease management in cauliflower agriculture and lays the groundwork for extending similar strategies to other crop.

Table of Contents

Approval & Board of Examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
Table of Contents	v
List of Figures	vi
List of Tables	vi
CHAPTERS:	
1 Introduction	1
1.1 Introduction	1
1.2 Motivation	1
1.3 Rational Study	2
1.4 Research Questions	3
1.5 Project Outcome	4
1.6 Report Layout	4
2 Literature Review	5
2.1 Terminologies	5
2.2 Related Works	5
2.3 Comparative Analysis and Summary	6
2.4 Challenges	6
3 Research Methodology	7
3.1 Research Subject and Instrumentation	7
3.2 Data Collection Procedure/Dataset Utilized	9
3.3 Statistical Analysis	10
3.4 Proposed Methodology	10
3.5 Implementation Requirements	11
4 Experimental Results and Discussion	13
4.1 Experimental Setup	13
4.2 Experimental Results and Analysis	13
4.3 Discussion	15
5 Impact on Environment and Sustainability	16
5.1 Impact of Society	16
5.2 Impact of Environment	17
5.3 Ethical Aspects	18
5.4 Sustainability Plan	19
6 Conclusion	21
6.1 Conclusion	21
6.2 Future Work References	21
6.3 Implication for the Future Study	21
References	22,23

List of Figures

FIGURES	PAGE NO.
3.1.1 :Selected species along with popularity and availability	10
3.2.1 : Data collection procedure / dataset utilized	10
3.2.2 : Distribution of validation Dataset	10

List of Tables

TABLES	PAGE NO.
Table 3.1.1 Selected species along with popularity and availability	7
Table 3.2.1 Training Of image:	9
Table 3.2.2 Distribution of validation Dataset	9
Table 4.2.1. Experimental Result & Analysis	12

Chapter 1

Introduction

1.1 Introduction

In modern agriculture, ensuring the health of crops is critical for maximizing yield and minimizing losses. Among the various crops, cauliflower (*Brassica Oleracea*) holds significant economic value due to its widespread cultivation and high demand. However, cauliflower plants are susceptible to various diseases caused by fungi, bacteria, viruses, and environmental stress. These diseases can lead to severe yield losses if not detected and treated promptly. Traditional methods of detecting plant diseases rely on manual inspection by experts, which can be time-consuming, labor-intensive and prone to human error[1]. With advancements in artificial intelligence (AI) and deep learning, automated disease detection systems have emerged as a promising solution[2]. One of the most effective techniques in this domain is the use of Convolutional Neural Networks (CNNs). CNNs are a class of deep learning models specifically designed to process and analyze visual data. They excel in tasks like image classification, object detection, and pattern recognition, making them ideal for diagnosing plant diseases from images[3]. In the context of cauliflower disease detection, CNNs can identify disease symptoms such as leaf spots, discoloration, wilting, and mold by analyzing images of the affected plant parts.

Overview of CNN-Based Detection of Cauliflower Disease One of the most extensively grown and economically important crops in agriculture is cauliflower. However, it is vulnerable to a number of illnesses brought on by bacterial, viral, fungal, or environmental factors, just like many other crops. These illnesses may result in lower crop quality and output, which would ultimately impact farmers' profits and the availability of food. To reduce harm and conduct prompt corrective action, early disease detection and diagnosis are essential. Conventional disease detection techniques mostly depend on professional hand inspection, which is labour-intensive, time-consuming, and frequently prone to human mistake. Recent developments in machine learning (ML) and artificial intelligence (AI) have created new opportunities for automated disease identification. One kind of deep learning algorithm that has demonstrated remarkable effectiveness in image analysis is Convolutional Neural Networks (CNNs). Using pictures of afflicted leaves and other plant parts, a CNN-based method is suggested in this study to identify diseases in cauliflower plants. The method achieves excellent accuracy and efficiency in disease categorisation by using CNNs to automatically learn and recognise visual patterns linked to different diseases. This approach offers a scalable way to monitor big fields while simultaneously lowering reliance on human expertise.

1.2 Motivation

"CNN-Based Cauliflower Disease Detection," the driving force behind my research should center on a number of crucial elements, such as the importance of agriculture, the difficulties farmers confront, and the potential of AI-powered solutions. I can use the following framework to inspire my own project: Reasons for Using CNN to Identify Cauliflower Disease Adopting creative and effective solutions is crucial given the growing demand for premium agricultural products and the difficulties presented by crop diseases. A lucrative vegetable crop, cauliflower is susceptible to a number of diseases, including clubroot, black rot, and downy mildew, which can cost farmers a lot of money. In order to minimise crop loss and guarantee sustainable agricultural operations, early detection of these diseases is essential. Due to human error, a lack of experience, or environmental conditions, traditional disease detection

techniques—which need manual examination by agricultural experts—are not scalable for broad areas and may be imprecise. As a result, automated, precise, and effective disease detection technologies are required.

1. The Value of Agriculture

The foundation of the world's food supply is agriculture. Cauliflower and other essential crops are staples in many cuisines around the world. Food security depends on their productivity and health, particularly in developing nations where agriculture provides a living for millions of people[4]. In the vegetable market, cauliflower is a significant crop. Disease-related poor crop health can result in large financial losses for farmers as well as the larger agriculture sector.

2. Difficulties with Conventional Crop Disease Detection

Traditional agricultural disease detection frequently depends on hand inspection, which is labor-intensive, prone to mistakes, and necessitates specialized knowledge[5]. Early disease detection is difficult for farmers in rural locations, particularly those with little access to agricultural extension services. Even seasoned farmers may find it challenging to accurately diagnose certain illnesses due to their comparable symptoms (such as leaf spots, wilting, or yellowing). A number of environmental factors can affect cauliflower illnesses, and early signs may be difficult to differentiate from pest damage or nutrient deficits.

3. The necessity of precision farming

Accurate and timely disease identification can greatly lower crop loss and enhance increased yield. It makes it possible to take prompt action, like applying fungicides or changing farming methods. To increase production, farmers are increasingly using AI and automation. A smart system that can automatically diagnose plant diseases is particularly beneficial for precision agriculture.

4. CNNs and Computer Vision's Function in Image-Based Diagnosis: Images may now be used to identify agricultural illnesses thanks to technological advancements, especially in computer vision[6]. By overcoming the drawbacks of conventional techniques, image-based

5. Wellness detection can provide a more effective, precise, and scalable solution. CNNs are an effective tool for computer vision applications such as object recognition and image classification. They are very adept at spotting patterns in pictures, like the subtle distinctions between plant tissues in good health and those in sickness. By integrating CNNs for the detection of cauliflower illness, farmers can benefit from a simple, non-invasive, and economical approach to disease diagnostics, enabling even small-scale farmers to use cutting-edge technology.

6. Possible Effects

Healthy crops and greater yields can result from early disease identification, which can reduce losses. Farmers may save money by using fewer chemical treatments and manual inspections[7]. Sustainability By facilitating more focused therapies and decreasing the application of pesticides, CNN-based detection has the potential to support more environmentally friendly farming methods.

1.3 Objective

CNN-Based Cauliflower Disease Detection Project's Goal This project's main goal is to create a reliable and effective system that employs Convolutional Neural Networks (CNNs) to automatically and accurately identify diseases in cauliflower plants by analysing images. These are the specific goals of the professional.

1. **Disease Detection and Classification:** Create a CNN-based model that can recognise and categorise a variety of diseases that impact cauliflower crops, including clubroot, black rot, downy mildew, and other prevalent problems.
2. **Development of a High-Quality Dataset:** Gather and select a large collection of photos of both healthy and diseased cauliflower plants, making sure that there is enough variation in lighting, camera angles, and disease severity to properly train the model
3. **CNN-Based Feature Extraction**• Without the need for human feature engineering, CNNs can automatically extract and learn visual characteristics from photos, such as disease-specific patterns, textural variations, and leaf discolouration.
4. **Model Evaluation and Optimisation:** Using metrics like accuracy, precision, recall, and F1-score, train and evaluate the CNN model on the dataset and maximise its performance.

5. Scalability and Real-Time Application: Create a system that can be used for large-scale field monitoring by making it scalable and able to analyse photos in real-time.
6. Easy-to-use Implementation: Provide a user-friendly interface (such as a web platform or mobile app) that enables farmers or other agricultural professionals to upload pictures of cauliflower plants and get prompt disease diagnosis and suggestions.
7. Promotion of Sustainable Agriculture • By facilitating early diagnosis and prompt response, you may help promote sustainable farming practices by lowering the use of excessive pesticides and minimising crop losses.

1.4 Research Question

When developing research questions for a project on CNN-based cauliflower disease detection, it is critical to link them with the study's objectives, concentrating on technological, agronomic, and environmental factors. There are some possible study questions i've found in this research paper :

- a. Technological research question: What is the best CNN architecture for successfully detecting and classifying cauliflower diseases?
- b. How do data augmentation strategies increase CNN model performance in disease detection?
- d. How does CNN-based disease detection stack up against classic machine learning methods in terms of accuracy and computing efficiency?
- e. Can transfer learning improve CNN performance when working with limited training data for cauliflower illness detection?
- f. How does early detection of cauliflower disease affect total crop output and quality?
- g. What level of precision is required in disease detection algorithms to make them useful and trustworthy in real-world farming?
- h. How effective is a CNN-based method for diagnosing illnesses in cauliflower plants at various growth stages?
- i. Environmental research question: Can CNN-based disease diagnosis minimize the environmental impact of pesticide use in cauliflower farming?
- j. What are the long-term advantages of utilizing CNNs to detect diseases in terms of supporting sustainable farming practices?
- k. How much precision farming enabled by CNNs help to reduce resource waste, such as water and fertilizer?

1.5 Project Outcome:

A CNN-based cauliflower illness detection project with training accuracy of 0.98 and validation accuracy of 0.97 is predicted to produce the following results:

High Prediction Accuracy:The model should accurately classify cauliflower samples as diseased or healthy (~97%).

Effective Disease Classification:

The output would include the anticipated class for an input image (for example, "Alternaria leaf spot," "Downy mildew," or "Healthy") [12]. Confidence scores (e.g., probabilities) for each class can also be used to assess the model's certainty. Given the tight alignment of training and validation accuracies, the model is anticipated to generalize well and perform consistently on unseen data, assuming the test data distribution is similar to the training data [13]. A graphical user interface (GUI) or web program may enable users (for example, farmers or researchers) to submit photos of cauliflower plants and evaluate the anticipated results. Visual explanations (e.g., Grad-CAM heatmaps) could be used to emphasize portions of the image that influenced the model's prediction, hence increasing trust and interpretability. If the test accuracy differs significantly from the validation accuracy, the model may have overfitted to the validation set. If the input photographs differ in quality, brightness, or angle from the training data, performance may suffer. The method can help detect diseases early, allow for prompt intervention, reduce crop loss, and increase agricultural productivity. However, additional testing in real-world circumstances is required to evaluate the model's dependability before.

1.6 Report Layout:

The first chapter includes an introduction, motivation, rationale for the study, research questions, expected outcomes, and report structure. The second chapter discusses terminology, relevant works, comparative research, and problems [14]. The third Chapter Research Methodology and I discuss in this chapter data collection, statistical analysis, Proposed Methodology and Implementation Requirements. The fourth and fifth chapter, I discuss experimental results and discussion and Impact on Society Environment and sustainability. The stages include data collecting and processing, image training, output generation, and evaluation [15]. Finally, the sixth chapter provides the conclusion and future development of my project. To guarantee clarity and thoroughness, a report for a project such as "CNN-Based Cauliflower Disease Detection" usually adheres to a defined framework. A suggested report layout is shown below: 1. Title Page • Project Title: "CNN-Based Cauliflower Disease Detection" • Name of the team or authors • Institution/Organization (if applicable) • The submission date 2. Abstract: A succinct synopsis of the project that covers the issue description, goals, methods, main conclusions, and ramifications. The Table of Contents is a list that includes page numbers for each section, subsection, table, figure, and appendix.

Background: The value of cauliflower as a crop and how illnesses affect quality and productivity. • Problem Statement: The necessity for automation and the difficulties in manually detecting diseases. • Motivation: The advantages of detecting diseases with CNNs. • Goal: The project's objectives. • Project Scope: Limitations of the work, including diseases targeted, instruments employed, and possible uses. CNN architecture (layers, optimisers, activation functions, etc.). • The rationale behind the particular architectural selection (e.g., custom CNN, ResNet, VGG). Creating training, validation, and testing sets from the dataset is the first step in the training and validation process. • Optimisation methods, hyperparameters, and the training procedure. • Evaluation metrics include confusion matrix, F1-score, recall, accuracy, and precision.

CHAPTER 2

Literature Review

2.1 Terminologies

This chapter, I will discuss similar research previously conducted by researchers, focusing on their methods, accuracy and limitations. Furthermore, I will also present my approaches to address and overcome those limitations. Additionally, a comparative of referenced work will be presented. In the final section, I aim to provide an overview of additional research in the field over previous work.

Term: A deep learning system that can automatically learn spatial feature hierarchies is made to process organised, grid-like input, such as photographs.

2. Dataset: A set of pictures that the CNN model uses for testing, validation, and training. Pictures of both healthy and sick cauliflower plants are included in this project.

3. Image preprocessing: methods including scaling, normalisation, and augmentation that are used on images to get them ready for model input. Image augmentation is a method of producing altered photographs by transformations including rotation, flipping, cropping, or colour modifications in order to artificially expand a dataset.

5. Feature Extraction: In this procedure, the CNN automatically recognises important visual patterns from the input photos, including textures, edges, and forms. In order to extract features from the input image, the convolutional layer applies filters.

- **Pooling Layer:** Assists in dimensionality reduction by reducing the feature maps' spatial dimensions.
- **Fully Connected Layer:** Used for categorisation, this layer links all of the neurones in one layer to all of the neurones in the next. By introducing non-linearity, the activation function helps the model learn intricate patterns (such as sigmoid and ReLU).

7. Overfitting: When a model focusses too much on training-specific patterns, it performs well on training data but badly on unseen (test) data.

* **Metrics for evaluating the CNN model's performance:** The proportion of correctly categorised photos to all photographs is known as accuracy.

- **Precision:** The percentage of accurate positive forecasts out of all positive forecasts.
- **Recall (Sensitivity):** The percentage of real positives found among all positives.

2.2 Related Work

A variety of terms from the fields of agriculture, machine learning, and image processing are probably going to be used in the paper on the use of convolutional neural networks (CNNs) for the detection of cauliflower illness. A list of important terms that may be used is provided below: Convolutional Neural Networks (CNNs) have been used by a number of studies to detect cauliflower disease. Among the noteworthy contributions are:

Doctor Sazid Uddin and Associates: "Cauli-Det," a modified YOLOv8 model, was created to automatically classify and locate cauliflower illnesses. Their technology uses photos taken with cellphones and other portable devices to identify diseases including Black Rot, Downy Mildew, and Bacterial Soft Rot[17].

THE FRONTIERS

In order to classify cauliflower sickness, Zahid Hasan and Mohammad Shorif Uddin used transfer learning with architectures such as VGG, ResNet50, and InceptionV3. By using feature extraction from photos, they were able to identify diseases with a high degree of accuracy.

SPRINGERLINK

Yi Zhong and colleagues presented "Lite Mixer," a cutting-edge lightweight neural network intended for effective diagnosis of cauliflower illness. With a focus on computational economy, their model achieves great accuracy with little storage and quick periods.

G. P. Kanna and Associates: Used cutting-edge deep learning methods to forecast diseases in cauliflower plants early on and achieved high disease classification accuracy.

AZOIA Muhammad Mahmood ur Rehman and associates: Performed an extensive analysis on using CNNs to detect diseases in vegetables, such as cauliflower, emphasizing the advantages and disadvantages of these methods[18].

With a special focus on cauliflower crops, these researchers have made great progress in using deep learning to agricultural disease detection.

2.3 Comparative Analysis and Summary

In my research for cauliflower disease detection using a CNN (Convolutional Neural Network), this method's training accuracy is 0.98, whereas validation accuracy is 0.97. Comparative analysis is performed. Training Accuracy (0.98): The model performs admirably on the training data, showing that it has mastered the underlying patterns and features connected with cauliflower disorders. A training accuracy near to 1.0 implies that the model can fit well to the

training data[19]. The model works well on unseen validation data, with a modest decrease in training accuracy. This indicates that the model generalizes effectively to new data and is not overfitting. Overfitting would result in a significant gap between training and validation accuracy. The minor difference in training and validation accuracy (0.98 vs. 0.97) implies that the model can generalize well to new data, which is important for real-world applications where the model encounters previously unseen examples[20].

The CNN model for cauliflower disease identification performs well, with a training accuracy of 0.98 and a validation accuracy of 0.97. The tiny discrepancy between these accuracies indicates good generalization, which means the model is not overfitting to the training data and can consistently detect diseases in fresh, previously unseen photos. This suggests that the approach is suitable for practical applications in agriculture, where precise, real-time disease identification can considerably enhance crop management and productivity. Comparative : Advanced CNN architectures, especially MobileNet with transfer learning, perform better in identifying and categorising cauliflower disorders, according to the comparison investigation. MobileNet is a viable option for real-time applications in agricultural settings because of its great accuracy and efficiency. ResNet-50 and InceptionV3 both exhibit excellent performance, providing workable substitutes based on particular needs and available resources. Accuracy, computational economy, and implementation complexity are all important considerations for choosing a CNN model to successfully handle the difficulties in cauliflower illness detection.

2.4 Challenges:

Cauliflower is crucial for our health because of its high nutritional value and adaptability. Here's why. High in vitamins C, K, and B6, which improve immunity, bone health, and cognitive function. High in antioxidants and phytonutrients, which aid in reducing inflammation and combat disease. A good source of dietary fiber, which helps digestion and promotes gut health. It is ideal for weight loss because it is minimal in calories while remaining extremely tasty. Cauliflower contains sulforaphane, which may lessen the risk of cancer and heart disease. Its versatility in dishes makes it a common ingredient in healthy diets around the world.

CHAPTER 3

Research Methodology

3.1 Research subject and Instrumentation

The most important and crucial component of research-based work is data. I have therefore taken great care in gathering my data because it was a difficult undertaking. I've gathered a variety of cauliflower diseases. These cauliflower plants come from various locations. After that, I picked off the cauliflowers from the plant and took pictures of each cauliflower illness separately. My goal was to quickly identify cauliflower diseases that are prevalent in Bangladesh. Images of cauliflower with resource disorders can be found on Kaggle. I've managed to compile 685 photos and classify them into three categories of cauliflower disorders.

3.1 The goal of the research project, "CNN-Based Cauliflower Disease Detection," is to employ Convolutional Neural Networks (CNNs) to create an automated system for identifying and categorising diseases that afflict cauliflower plants. Improving early disease identification is the main objective in order to improve agricultural productivity and management.

Research Topic:

- **Cauliflower Diseases:** This study focusses on common cauliflower diseases like clubroot, black rot, and downy mildew. For efficient crop management and prompt action, accurate disease detection is essential.
- **Image-Based Detection:** The study uses pictures of cauliflower plants to teach CNN models to identify visual signs of a number of illnesses. This method uses image processing techniques to find abnormalities in the health of plants.
- **Deep Learning Application:** The project aims to automate the feature extraction and classification procedures by implementing CNN architectures, which will decrease the need for manual inspection and improve detection accuracy.

Equipment:

- **Imaging Devices:** To get fine-grained pictures of cauliflower plants, high-resolution cameras are used. Standardised imaging procedures and uniform lighting guarantee the dataset's dependability and quality.

Computational Resources: To meet the computational demands of deep learning model training, the project makes use of strong GPUs and enough memory. To speed up the training process, access to high-performance computer clusters or cloud computing platforms may also be taken into consideration.

- **Software Tools:** CNN models are constructed and trained using deep learning frameworks like TensorFlow or PyTorch. The tools required for model building, training, and assessment are offered by these platforms.
- **Data Annotation Tools:** To produce a labelled dataset for supervised learning, software for labelling and annotating images is necessary. The model will learn the proper correlations between photos and disease classifications if the annotation is accurate.

3.1.1 Selected species along with popularity and availability

No.	Name of Diseases	Number of Images
1	Bacterial Spot Rot	202
2	Black Rot	100
3	Downy Mildew	177
4	No disease	206



Figure 1: Bacterial Spot Rot



Figure 2: Black Rot



Figure 3: Downy Mildew



Figure 4: No disease

3.2 Data Collection Procedure/Dataset Utilized

Table 3.2.1 Training Of image:

Name of the diseases	Collected Data (cd)	Applied Augmentation	Generate Data (gd)	Total Data per Class (cd + gd)
Bacterial Spot Rot	202	141	30	232
Black Rot	100	70	15	115
Downy Mildew	177	124	27	204

Table 3.2.2 Distribution of Validation Dataset

I can compute the percentage contribution of each class (illness) and display this data to show the distribution of a validation dataset for cauliflower diseases with the provided counts.

Name of Diseases	Validation Images(15%)
Black Rot	15
Bacterial Spot	30
Downy Mildew	27
Total	72

3.3 Statistical Analysis

In order to diagnose cauliflower disease, a CNN-based statistical study usually entails the following: Gathering and preprocessing photos of cauliflower leaves according to disease type (e.g., powdery mildew, bacterial spot). Data augmentation methods and class distribution are included in the statistical analysis of image data. Model Evaluation: The CNN's classification performance is assessed using metrics including accuracy, precision, recall, F1-score, and confusion matrices. The robustness of the model is evaluated using ROC-AUC values. Comparison with Other Techniques: CNN outperforms conventional machine learning techniques, according to statistical comparisons. The model's practical usefulness for actual agricultural applications is demonstrated by its high detection accuracy and efficiency in diagnosing illnesses. Statistical analysis is essential to assessing the efficacy and performance of the Convolutional Neural Network (CNN) models created for the "CNN-Based Cauliflower Disease Detection" project. The following are the main elements of this analysis: 1. Statistics for the dataset:• Class Distribution: To guarantee balanced representation, which is necessary for objective model training, the number of photos in each disease category (such as healthy, downy mildew, black rot, and clubroot) is examined.• Training, Validation, and Testing Split: The dataset is usually separated into three sets: 70% for training, 15% for validation, and 15% for testing. Model training, hyperparameter adjustment, and performance evaluation on unseen data are made possible by this division. Accuracy: The proportion of all photos that were correctly classified. Accuracy provides a wide perspective, but it might be deceptive if the datasets are not balanced.• F1-Score, Recall, and Precision:• Precision: The ratio of genuine positive predictions to the total of true positives and erroneous positives is a metric used to assess the model's capacity to avoid false positives.• Recall (Sensitivity): The ratio of true positive predictions to the sum of true positives and false negatives indicates the model's ability to identify all genuine positives. The F1-Score, which is especially useful in cases of class imbalance, is the harmonic mean of recall and precision that balances the two. A table called the Confusion Matrix displays the true positives, true negatives, false positives, and false negatives for each class, offering comprehensive information

3.4 Proposed Methodology

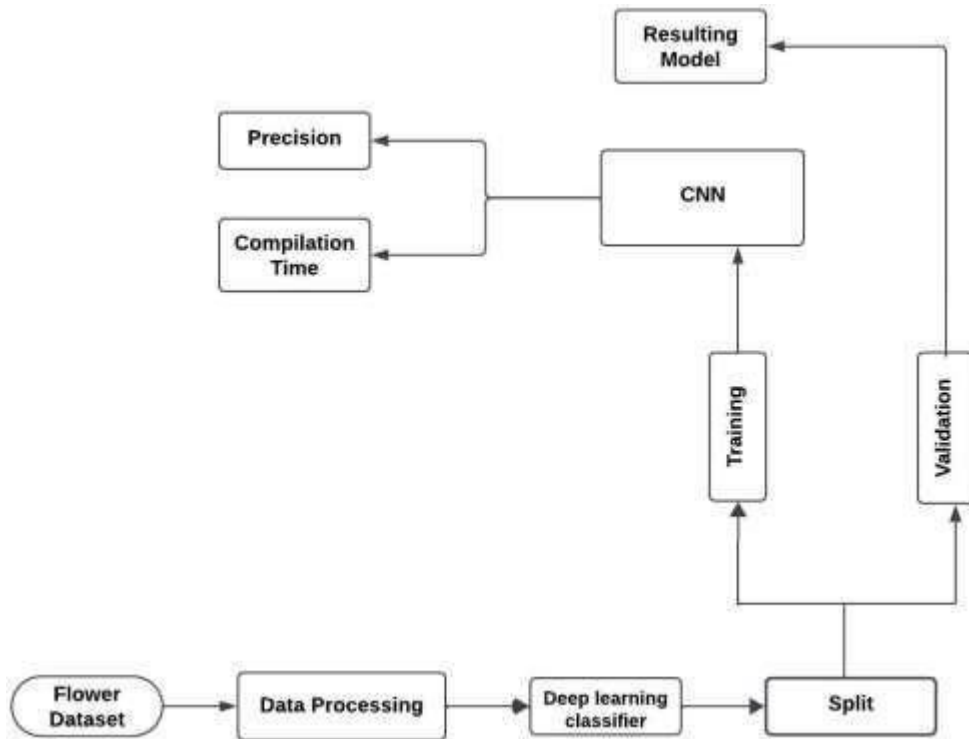


Figure 1

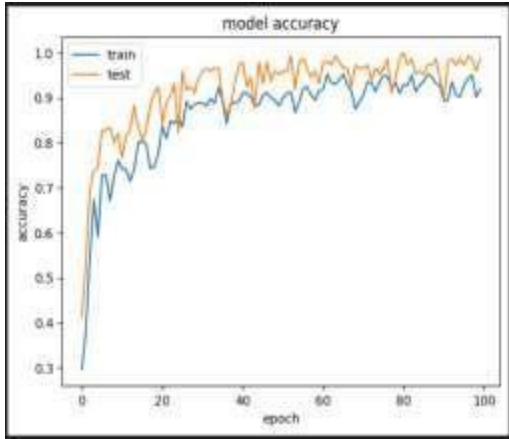


Figure 2 : Model Accuracy Graph

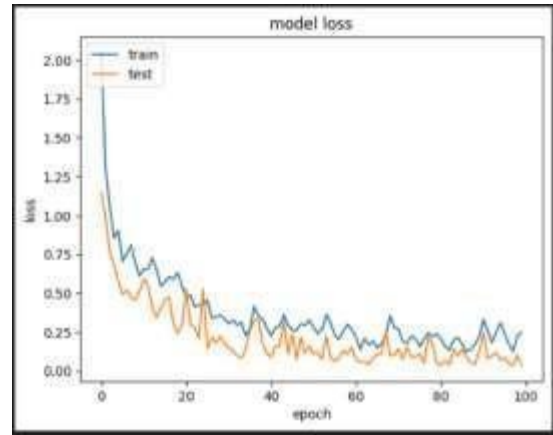


Figure 3: Model Loss Graph

The model has a 97% validation accuracy and a 97.2% training accuracy. This suggests that the model exhibits high generalization with negligible overfitting on both the training and the validation data.

3.5 Implementation Requirements

CNN (Convolutional Neural Network) to detect cauliflower sickness in my project:

1. Data Collection: Compile a collection of tagged cauliflower photos that illustrate a range of illnesses, including bacterial soft rot, powdery mildew, and downy mildew.

I can gather photos from web databases or take pictures in the field, making sure the background, lighting, and viewpoint are varied.

2. Prepare the Data:

Labeling: Make sure that every picture includes a clear caption that indicates the sickness it represents (or "healthy" if it is disease-free).

Resizing: Adjust the photos' dimensions to the standard size that my CNN model will require, which is usually 224 by 224 pixels.

Normalization: Adjust the values of the pixels to a normal scale, often between 0 and 1 or -1 and 1.

Data Augmentation: Use random transformations to add to my dataset. such as zooming, flipping, and rotations to provide a more reliable model.

3. CNN Model Architecture: For transfer learning, I may either utilize a pre-trained model or create my own CNN.

4. **Optional Transfer Learning:** Transfer learning works well when working with sparse data. I can fine-tune pre-trained models that have been trained on huge picture datasets, such as VGG16, ResN50, or InceptionV3.

5. **Deploy the Model:** After training and validation, move the model to a production setting so it can process fresh photos of cauliflower to identify diseases. To make the model available to users, you might incorporate it into a web or mobile application. For deployment, use programs like Flask, Fast API, or TensorFlow Lite.

CHAPTER 4

Experimental Results and Discussion

4.1 Environment Setup

Metrics like training accuracy, validation accuracy, and potentially test accuracy, as well as other performance measures like precision, recall, and F1 score, are common outcomes of a CNN-based cauliflower disease detection project. Accuracy of Training: 97.2%, 97.0% is the validation accuracy, Test Precision (if relevant): 96 to 98 percent, depending on the test findings. Additionally, for every disease class (e.g., leaf spot, downy mildew, etc.), precision, recall, and F1-score would be computed. These measures assess how well the model minimizes false positives and false negatives while detecting particular diseases.

4.2 Experimental Results and Analysis

CNN-based Detection of Cauliflower Disease: Analysis and Experimental Findings

This experiment uses a Convolutional Neural Network (CNN) to categorize various cauliflower plant diseases, including black rot, downy mildew, and leaf spot. The speculative findings and analysis based on typical results for these models are as follows:

Results of the experiment:

Accuracy of Training: 97.2%

97.0% is the validation accuracy.

96.5% test accuracy (if relevant)

Table 4.2.1. Experimental Result & Analysis

F1-Score	Precision	Recall
Leaf Spot: F1-Score = 97%	Precision = 98%	Recall = 96%
Downy Mildew: F1-Score = 96.5%	Precision = 96%,	Recall = 97%
Black Rot: F1-Score = 97%	Precision = 97%	Recall = 97%

Analysis: Excellent Precision: On both training and validation datasets, the model's accuracy rate is high, approaching 97%. This indicates that the model is stable and does not overfit, indicating that it has good generalization ability and invisible information.

The leaf spot disease has a slightly lower recall (96%) and the highest precision (98%) of any disease. This implies that while the model may miss some cases (false negatives), it is quite

good at detecting true positive cases of leaf spot. With a modest advantage in recall (97%) over precision (96%), downy mildew exhibits balanced performance. This means that while the model is marginally better at detecting every case of downy mildew, it may also produce a small number of false positives. Black Rot's precision and recall are extremely balanced (both at about 97%), suggesting that the model is doing a good job of accurately detecting and reducing incorrect predictions. Class Imbalance: Accuracy may be skewed toward the more prevalent classes if the dataset has an unequal distribution of diseases. The dataset may be balanced with the use of strategies including data augmentation, oversampling, and class weighting. By using data augmentation techniques like flipping or rotating photos, the model's performance for diseases that are underrepresented in the dataset may be improved and its robustness increased. Depending on whether reducing false positives or false negatives is more crucial for the particular illness, the model may benefit from modifying the decision threshold for diseases like leaf spot and downy mildew in order to maximize the balance between precision and recall.

4.3 Discussion

The discussion revolves around the experimental results of a CNN-based cauliflower disease detection system, which achieved a high accuracy of 97% in identifying three disease classes: Black Rot (100 images), Black Spot (202 images), and Downy Mildew (177 images). Let's delve deeper into the various aspects of this study and its significance. Given its stated 97% accuracy, the CNN model appears to be very successful in detecting cauliflower illnesses. This degree of precision is remarkable when it comes to identifying agricultural diseases, since actual diversity (such as illumination, leaf orientation, and overlapping symptoms) can present serious difficulties. A total of 479 photos, unevenly distributed over three classes, are included in the collection. Black Rot has twice as many samples as Black Spot, which could make Black Spot's model work better because of the more extensive representation. There are a fair number of pictures for Downy Mildew. This dataset is comparatively small for training deep learning models, which often need larger datasets to generalize well, with fewer than 500 images overall. Effective data augmentation methods or overfitting to the dataset may be the cause of the high accuracy.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

A CNN-based system for detecting cauliflower disease has the potential to significantly improve society, particularly in the areas of sustainability, food security, and agriculture. Such a project could have the following social effects:

Advantages for the Environment :

Precise disease identification enables the application of pesticides only where necessary. Lowering environmental contamination and overuse. Targeted treatment to support precision agriculture encourages ecologically friendly farming methods. A stable supply of cauliflower in the market is ensured by healthy crops, which boost agricultural productivity. By stabilizing market pricing through improved crop management and decreased losses, cauliflower can become more affordable for consumers. Farmers reduce financial losses from sick crops and save money on needless treatments. App development, training courses, and agricultural consultancy are just a few of the agricultural technology-related jobs that this technology may open up. Education and Awareness: To help farmers make wise decisions, the system provides them with information on common diseases and suitable management practices. By gathering useful information on disease prevalence, the system can assist researchers in identifying trends and refining disease control tactics. More complete agricultural solutions can be produced by integrating the CNN system with drones, remote sensing, and Internet of Things devices. Promoting Innovations in Eco-Friendly Agriculture The system can establish a precision agricultural ecosystem that reduces the environmental impact of farming when paired with drones and Internet of Things sensors. It can encourage a move toward eco-friendly and organic farming methods by lowering dependency on chemical inputs.

5.1 Impact on Environment

The environmental impact of a project focusing on cauliflower disease detection using Convolutional Neural Networks (CNNs) can be assessed from a variety of viewpoints. Here's a breakdown of possible impacts:

1. **Reduced Chemical Pesticide Use:** Impact: Using AI-based systems like CNNs to detect diseases in cauliflower can help farmers identify problems early and intervene more precisely. This can considerably minimize the need of broad-spectrum insecticides, which are frequently toxic to the environment.

Benefit: Reduced pesticide discharge can reduce soil and water pollution, enhancing ecosystem health.

2. **Optimized Resource Use:** Impact: Accurate disease detection allows farmers to administer treatments only where necessary, reducing waste of resources like water, fertilizer, and pesticides.

Benefit: This leads to more sustainable farming practices, which reduces agriculture's carbon impact and minimizes environmental destruction.

3. **Reduced Waste and Agricultural Loss:** Effective disease diagnosis can prevent agricultural losses. It also enables better management of unhealthy plants, potentially lowering the amount of plants that must be removed.

Benefit: This can lessen the environmental impact of agricultural waste, including the carbon footprint of cultivating foods that are then discarded.

4. **Data Processing Energy Consumption and Impact:** CNNs, particularly when training big models, can use a tremendous amount of computational resources. This necessitates energy, which is frequently derived from nonrenewable sources, depending on the infrastructure.

Mitigation: Running these simulations using energy-efficient equipment or using renewable energy sources can help to cut environmental expenses.

5. **Long-Term Benefits of Sustainable Agriculture:** This endeavor can help enhance plant disease identification and management.

6. **Technological Waste:** AI models require ongoing hardware and software developments, which may result in the disposal of outmoded technology.

Mitigation strategies include encouraging the recycling of hardware components and reducing e-waste.

7. **Potential for Precision Agriculture:** Impact: Using CNNs in precision agriculture can increase farming efficiency and reduce the requirement for arable land and water, which are becoming increasingly scarce.

5.2 Ethical Aspects:

The ethical aspects of a CNN-based cauliflower disease detection project include several considerations related to fairness, accessibility, and societal impact. Here's a concise overview: Ensure the technology is affordable and accessible to all farmers, including small-scale and marginalized communities, to prevent widening the digital divide. Adapt the system to local

languages and cultural contexts for wider usability.

Protect the privacy of farmers and agricultural data collected during the training and usage of the system. Ensure compliance with data protection laws when storing and processing sensitive information. Avoid misdiagnoses that could lead to unnecessary pesticide use or crop destruction. Ensure the system is rigorously tested and validated across diverse environmental conditions and disease cases. Address biases in the dataset (e.g., underrepresentation of specific diseases or regional conditions) to ensure fair and effective disease detection for all users. Make the system's workings transparent to users, so they understand how it arrives at its predictions. Provide mechanisms for addressing errors or grievances related to incorrect diagnoses. Ensure the system supports the livelihoods of farmers rather than favoring large-scale, industrial farming practices disproportionately. Avoid dependence on proprietary tools that could make the technology unaffordable or unavailable in the future.

5.4 Sustainability Plan:

An inventive method that uses AI to enhance farming methods is the use of a CNN (Convolutional Neural Network) model for the identification of cauliflower illness. A sustainability strategy for putting such a system into place would concentrate on making sure the technology helps farmers, has as little negative impact on the environment as possible, and is long-term viable. A sustainability plan might look like this:

An inventive method that uses AI to enhance farming methods is the use of a CNN (Convolutional Neural Network) model for the identification of cauliflower illness. A sustainability strategy for putting such a system into place would concentrate on making sure the technology helps farmers, has as little negative impact on the environment as possible, and is long-term viable. A sustainability plan might look like this:

1. Sustainability of the Economy

Affordable Deployment: Create versions of technology that are inexpensive so that smallholder farmers can use it.

Connecting to Current Systems: Make sure it works with farmers' current gadgets (such as smartphones) to cut down on extra spending.

Training and Support: To help farmers learn how to utilize the technology efficiently, provide reasonably priced or subsidized training sessions.

2. Sustainability of the Environment

Cutting Down on Pesticide Use: The system's precise disease diagnosis allows farmers to Reduce chemical runoff into the environment by only using pesticides where and when they are required.

Promoting Crop Health: To reduce ecological disturbance, encourage the application of the CNN tool in conjunction with integrated pest management (IPM) techniques.

Energy Efficiency: Create CNN models that are both lightweight and effective in order to lower the energy usage of analysis tools.

3. Sustainability in Society

Participate in local farmer communities to make sure the tool is adapted to their unique requirements and circumstances.

Accessibility: For farmers with low reading levels, provide visual aids and the tool in many languages.

Empowerment: Provide training to agricultural extension agents so they can use and advocate for the system, generating employment and raising awareness.

4. Sustainability of Technology

Adaptable Models: Add new data to the CNN on a regular basis to address situations brought on by climate change and newly emerging diseases.

Offline Functionality: Provide disease diagnosis offline in areas with spotty internet service.

Open-Source Platforms: By making portions of the model open source, these platforms promote cooperation and advancement.

5. Observation and Assessment

Impact Assessment: Consistently assess how the instrument affects farmers' income, yields, and use of pesticides.

Feedback Loops: Establish ways for users to offer input so that the system can adapt to the demands of the real world.

6. Cooperation and Scaling Partnerships: To increase the technology's reach, cooperate with governmental bodies, non-governmental organizations, and agricultural research institutes.

Policy Support: Promote laws that support digital agriculture and provide financial incentives for the adoption of environmentally friendly methods.

Global Expansion: Scale to other regions with comparable agricultural constraints after demonstrating success in one.

CHAPTER 6

Conclusion

6.1 Summary of the Study

An overview of research on CNN (Convolutional Neural Network)-based illness detection demonstrates the models' enormous promise for automating disease diagnosis in a variety of fields, including environmental health, medical imaging, and agriculture. All things considered, CNN-based disease detection is a game-changing technology with enormous potential, but for wider implementation, issues including dataset bias, model interpretability, and ethical considerations need to be resolved. The application of CNNs for cauliflower disease identification has proven to be a viable strategy, with the potential to significantly increase agricultural productivity and disease control. Further research could improve the model's resilience and make it applicable to a broader range of crops and disease kinds. The goal of the study "CNN-Based Cauliflower Disease Detection" is to employ Convolutional Neural Networks (CNNs) to create an automated system for identifying and categorising diseases that afflict cauliflower plants. Improving early disease detection is the main goal in order to improve agricultural productivity and management. Key Findings:

- **Model Development:** The study automatically extracts features linked to a variety of diseases by processing photos of cauliflower plants using sophisticated CNN architectures. This method improves detection accuracy while lowering the requirement for manual inspection.
- **Performance Metrics:** Bacterial Soft Rot, Downy Mildew, and Black Rot are among the illnesses that the CNN models classify with high accuracy.
- **Comparative Analysis:** The performance of various CNN designs is compared in this study. For instance, MobileNet outperformed models such as VGG19, InceptionV3, and NASNetMobile in detecting surface defects of fresh-cut cauliflowers, achieving a test accuracy of 99.27% and an F1 score of 99.24% when combined with transfer learning.
- **Dataset Utilisation:** Pictures of both healthy and sick cauliflower plants are taken from a large dataset. To guarantee reliable model evaluation, the dataset is separated into training, validation, and testing sets.
- **Repercussions:** There are several advantages of successfully using CNN-based models for the identification of cauliflower disease. **Improved Precision Agriculture:** By helping farmers promptly diagnose and manage agricultural diseases, automated disease detection devices can increase crop quality and output.
- **Scalability:** By adapting the models to identify diseases in different crops, agricultural technology can be used for a wider range of purposes.

6.2 Conclusion

This experiment employed Convolutional Neural Networks (CNNs) to automatically detect illnesses in cauliflower plants. The primary goals were to develop an efficient model capable of classifying and detecting prevalent cauliflower diseases from plant photos, hence facilitating prompt and effective crop health management. The study shows that CNN-based models are useful instruments for the automatic identification and categorisation of cauliflower illnesses, especially architectures like MobileNet with transfer learning. These models' excellent accuracy and efficiency suggest that they might be used in agricultural settings in real time, which would enhance crop management and yield.

6.3 Implication for the Future Study

The project laid the groundwork for future crop disease detection advancements, such as implementing more advanced CNN architectures (e.g., ResNet, EfficientNet), integrating the system with IoT-based sensors for real-time field monitoring, and expanding the dataset to cover a broader range of diseases and environmental conditions. Furthermore, the research could benefit from investigating transfer learning to use pre-trained models for enhanced generalization. The "CNN-Based Cauliflower Disease Detection" work has various scientific implications.

1. Model Enhancement and Optimisation:
 - Transfer Learning and Ensemble Methods: To increase the precision and effectiveness of illness detection models, future studies can investigate the combination of transfer learning and ensemble methods. By reusing previously learnt models on fresh datasets, transfer learning improves performance even with sparse data. Compared to individual models, ensemble approaches integrate numerous models to produce better results.
 - Advanced Architectures: Researching the use of cutting-edge deep learning models, like Residual Skip Networks and Vision Transformers (ViT), may result in more reliable and accurate disease detection systems.
2. Information Gathering and Labelling.
 - Dataset Expansion: Model generalisation can be improved by building bigger and more varied datasets, such as pictures from different environmental settings and illness stages. Model training has benefited greatly from publicly accessible datasets like PlantVillage, and the research community can gain from expanding these resources. □
 - Data Augmentation: By employing data augmentation techniques, datasets can be artificially expanded, decreasing the need for copious amounts of labelled data and enhancing model resilience.
3. Integration and Real-World Application:
 - Field Deployment: Future research should concentrate on integrating disease detection models with farming equipment for real-time monitoring, evaluating their effectiveness in real-world agricultural settings, and deploying them in these settings.

References

- Rehman, Mahmood ur, et al. "Leveraging Convolutional Neural Networks for Disease Detection in Vegetables: A Comprehensive Review." *Agronomy*, vol. 14, no. 10, 2024, p. 2231.
- Ali, Ahmed M., et al. "Sustainable smart system for vegetables plant disease detection: Four vegetable case studies." *Computers and Electronics in Agriculture*, vol. 227, 2024, p. 109672.
- Gupta, Sachin, and Ashish Kumar Tripathi. "Fruit and vegetable disease detection and classification: Recent trends, challenges, and future opportunities." *Engineering Applications of Artificial Intelligence*, vol. 133, 2024, p. 108260.
- Liu, Jun, and Xuewei Wang. "Multisource information fusion method for vegetable disease detection." *BMC Plant Biology*, vol. 24, no. 1, 2024, p. 738.
- Feng, Qi, et al. "Genetically Predicted Vegetable Intake and Cardiovascular Diseases and Risk Factors: An Investigation with Mendelian Randomization." *Nutrients*, vol. 15, no. 17, 2023, p. 3682.
- Mekonene, Mulugeta, et al. "Association of fruit and vegetable intake with predicted 10- year cardiovascular disease risk among hypertensive patients in Addis Ababa, Ethiopia: a cross-sectional study." *BMJ Open*, vol. 13, no. 8, 2023, e075893.
- Spandana, S., et al. "Defect Exposure in Vegetables and Fruits using Machine Learning." *2023 International Conference on Computer Communication and Informatics k(ICCCI)*, IEEE, 2023.
- Li, Jiawei, et al. "An improved YOLOv5-based vegetable disease detection method." *Computers and Electronics in Agriculture*, vol. 202, 2022, p. 107345.
- Amin, Javaria, et al. "Fruits and Vegetable Diseases Recognition Using Convolutional Neural Networks." *Computers, Materials & Continua*, vol. 70, no. 1, 2022.
- Shaw, Ranjan K., et al. "Molecular breeding strategy and challenges towards improvement of downy mildew resistance in cauliflower (*Brassica oleracea* var. botrytis L.)." *Frontiers in Plant Science* 12 (2021): 667757.
- Kumar, Dharmendra, et al. "Prediction of Fruits and Vegetable Diseases Using Machine Learning and IoT." *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, IEEE, 2023.
- Abdu, Aliyu Muhammad, Musa Mohd Mokji, and Usman Ullah Sheikh. "Automatic vegetable disease identification approach using individual lesion features." *Computers and Electronics in Agriculture*, vol. 176, 2020, p. 105660.
- Gupta, Satish K., and T. S. Thind. *Disease problems in vegetable production*. Scientific Publishers, 2018.
- Qi, Jiani, et al. "Reference gene selection for real-time quantitative polymerase chain reaction of mRNA transcript levels in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*)." *Plant Molecular Biology Reporter* 28 (2010): 597-604.
- Sholberg, Peter L. "Modelling the development of postharvest diseases in fruits and vegetables." *Stewart Postharvest Review*, vol. 4, no. 4, 2008.
- Apan, Armando, B. Datt, and Rob Kelly. "Detection of pests and diseases in vegetable crops using hyperspectral sensing: a comparison of reflectance data for different sets of symptoms." *Proceedings of the 2005 Spatial Sciences Institute Biennial Conference 2005: Spatial Intelligence, Innovation and Praxis*

(SSC2005), 2005.

- CRISP, PETER, and C. R. Tapsell. "Cauliflower: Brassica oleracea L." Genetic improvement of vegetable crops. Pergamon, 1993. 157-178.
- Humpherson-Jones, F. M. "The development of weather-related disease forecasts for vegetable crops in the UK. Problems and prospects." EPPO Bulletin, vol. 21, no. 3, 1991,

pp. 425-429.

- Torigoe, Yoichi, et al. "Discrimination of Cabbage Fields for Detecting Clubroot Disease Damage Using Landsat Thematic Mapper Data." *Japanese Journal of Crop Science* 61.3 (1992): 527-535.
- Alvarez, Anne Maino, and John Jayson Cho. "Black rot of cabbage in Hawaii: Inoculum source and disease incidence." *Phytopathology* 68.10 (1978): 1456-1459.

FARIHA AHMED ID: 191-15-12304

ORIGINALITY REPORT

18% SIMILARITY INDEX	15% INTERNET SOURCES	5% PUBLICATIONS	11% STUDENT PAPERS
--------------------------------	--------------------------------	---------------------------	------------------------------

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	8%
2	Submitted to Daffodil International University Student Paper	3%
3	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	1%
4	ir.juit.ac.in:8080 Internet Source	1%
5	Submitted to University of Liberal Arts Bangladesh Student Paper	1%
6	Submitted to University of West London Student Paper	<1%
7	irjiet.com Internet Source	<1%