Graph Neural Network Based Framework for Accurate Crop Disease Identification.

By

Md. Mostafizur Rahman

201-15-13849

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised by

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

Co-Supervised by

Md. Shahriar Shakil
Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

January 2025

APPROVAL

This Project titled **"Graph Neural Network Based Framework for Accurate Crop Disease Identification"**, submitted by Md. Mostafizur Rahman, ID No: **201-15-13849** 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 12th January, 2025.

BOARD OF EXAMINERS

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

Dr. Md Alamgir Kabir (DMAK) **Internal Examiner 1**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Tanvirul Islam (TI) **Internal Examiner 2**
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Manowarul Islam **External Examiner**
Associate Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. Md. Zahid Hasan, Associate Professor,** Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree.

Supervised by:

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

Co-Supervised by:

Md. Shahriar Shakil
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

12.01.25

Md. Mostafizur Rahman
Student ID: 201-15-13849
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. We are deeply grateful to everyone who has assisted us 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 **Dr. Md. Zahid Hasan**, **Associate Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **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.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

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

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

ABSTRACT

Crop diseases are a major challenge to agricultural productivity and hence diseases it at is the crucial earliest to identify to and avoid detect huge the loss of crops and ensure approach food for security. identifying the crop study diseases presents through a a new machine learning model in the (GNN). form the of work Graph is Neural divided Network into six phases to accomplish the paper's objective, collecting the a first large of dataset which from entails a public domain that comprises images of crop diseases Powdery Mildew including and Ash Bitter Gourd Gourd Downy Mildew. These images are further divided into different disease categories to facilitate specific and examination. Further, employed several on image the processing techniques. In (GLCM) the with feature statistical extraction measures also with Co-occurrence deep Matrix learning extracted features using DenseNet121. Both kinds of features are critical for disease identification as as well they as provide high-level patterns. both detailed The most important contribution of this CropGNN study model is which the implements proposed graphs to represent the correlations between the features that have provides been an extracted. effective way for This Crop framework Disease Detection graph-based and learning. Through it enhancing integrates the state-of-the-art disease machine identification, learning this methods approach with enables better decision making in agriculture and helps in addressing one of the biggest challenges of our time which is sustainable farming and food security.

Table of Contents

Approval i

Declaration ii

Acknowledgements iii

Abstract iv

List of Figures vii

List of Tables viii

1 Introduction 1

1.1 Introduction 1
1.2 Motivation 2
1.3 Objectives 3
1.4 Methodology 3
1.5 Project Outcome 4
1.6 Organization of the Report 5

2 Background 7

2.1 Introduction 7
2.2 Literature Review 8
2.2.1 Similar Work 16
2.2.2 Related Research 17
2.3 Gap Analysis 17
2.4 Summary 19

3 Research Methodology 20

3.1 Methodology 20
3.1.1 Overview 20
3.2 Detailed Methodology 21
3.2.1 Dataset Description 21
3.2.2 Data Pre-Processing 22

3.2.3 Feature Extraction..2
3.3 Architecture of CropGNN Model 23
3.4 Project Plan..24
3.5 Task Allocation..24
3.6 Summary 25

4 Implementation and Results 26
4.1 Environment Setup 26
4.2 Performance Metrics 26
4.3 Ablation Table 27
4.4 Comparative Analysis 28
4.5 Results and Discussion 29
4.6 Summary 32

5 Engineering Standards and Design Challenges 33
5.1 Impact on Society, Environment and Sustainability 33
5.1.1 Impact on Life.. 33
5.1.2 Impact on Society & Environment.. 33
5.1.3 Ethical Aspects 33
5.1.4 Sustainability Plan..33
5.2 Project Management and Financial Analysis......................................34
5.3 Complex Engineering Problem.. 34
5.3.1 Complex Problem Solving... 34
5.3.2 Engineering Activities.. 34
5.4 Summary 36

6 Conclusion 37
6.1 Summary 37
6.2 Limitation 37
6.3 Future Work 38

References 39

List of Figures

3.1.1 Main workflow of the proposed work 21

3.3.1 Architecture of CropGNN Model 23

4.5.2.1 Confusion Matrix 30

4.5.3.1 ROC Curve 31

4.5.4.1 t-SNE Node Embedding 31

3.1.1 Main workflow of the proposed work 21

List of Tables

2.2.1 Comparative analysis with previous work.. 11
3.2.1.1 Number of images corresponding to the class. 21
4.3.1 Ablation Table 27
4.5.1.1 Performance Metrics Table.. 29
5.2.1 Cost Estimate Table 34
5.3.1.1 Mapping with complex problem solving. 34
5.3.1.2 Mapping with knowledge Profile. 35
5.4.2.1 Mapping with complex engineering activities. 36

Chapter 1

Introduction

Bangladesh, with its agriculture-driven economy, faces significant challenges due to plant diseases affecting crop yields. This study proposes an innovative machine vision-based system for early disease detection in gourds, using machine learning to enhance agricultural productivity.

1.1 Introduction

Agriculture holds the central core in socio-economic aspects among nations, generally those which are rural or agriculture-dominated, like Bangladesh. Such as, approximately 66% of the total population still resides in villages where agriculture is the key source for their livelihoods, providing job opportunities to close to 60% of the available workforce with approximately a 12.46 percent contribution to GDP [1]. The agricultural sector in Bangladesh faces a number of challenges, major among them being plant diseases, despite the sector's importance. The plant diseases seriously threaten crop yield and quality. This bears great significance not only in food security but also in economic stability, since crop losses due to diseases result in significant yield loss and financial losses among farmers [2].

Most plant pathogens like bacteria, fungi and viruses have been causing diseases and rendering crops damaged worldwide. Economically, significant losses arise from these, hence contributing to food insecurity problems around the globe, while losses run over USD 220 billion annually [3]. Climate change, that contributes to increased disease outbreaks of different crops in monoculture plantations with limited genetic diversities, further deteriorates this situation in Bangladesh [4]. Crops such as tomatoes, cassava, strawberries, and gourds (ash gourd, bitter gourd, bottle gourd, snake gourd) are vulnerable to powdery mildew, downy mildew, and other leaf spots. These diseases affect not only agricultural yield but also the nutritional value of crops, thus adding to food shortages [5].

The increasing menace of plant diseases demands serious attention in the development of disease management approaches. Conventional farming is still in practice in rural Bangladesh, where traditional methods don't help much to stop this rapid spread. Innovative approaches will be more crucial through machine learning and image processing. Early detection through technology can support minimizing crop losses, hence helping to improve food security with timely interventions.

This research paper tries to bridge the gap between traditional farming and new, advanced technology by designing a mobile-based agro-medical expert system that makes use of deep learning models, especially CNNs, in identifying diseases in crop images captured from mobile devices. Advanced preprocessing techniques and extraction methods are used for feature processing to provide real-time, highly accurate disease detection, thereby enabling farmers to take the right decisions It is two-pronged: first, it's a problem of a lack of awareness and resources regarding disease control, second, and very importantly, there has to be the development of an actionable tool capable of practically determining crop diseases to bring about effective management. The current paper tries to offer a low-cost and accessible solution for disease detection, which may improve crop yields and enhance food security so that poor farmers are able to assure livelihoods.

1.2 Motivation

This is motivated by the increasing challenges that the agricultural sector is facing due to the devastating impact brought about by the diseases of plants. Such a challenge is very critical, with countries like Bangladesh relying so heavily on agriculture for the sustenance of millions. In particular, plant diseases threaten crop yield and quality; food security and economic stability, based on farming; and also, generally affect the well-being of farm families. The stakes are bigger than ever with a growing population and food demand, at a time when resources might be low in some countries and when many farmers use traditional approaches.

Climate change and farming monoculture have only augmented the rate of plant diseases to an alarming rate, calling for urgent need for detection as early as possible. The solution to this problem can bring in transformative dividends in the form of reduced crop losses and an empowered farmer class with knowledge and tools to fight these challenges. Based on the current state of machine learning and image processing, this research aims at developing a robust, accessible, and low-cost early plant disease detection system. Outcomes of this study may indeed advance sustainable agriculture, ensure more dignified livelihoods among farming families, and contribute to improved global food security for hopes of a healthier and stronger future.

1.3 Objectives

Therefore, the objectives of this research are stipulated below.

- **Enhancing Early Detection of Plant Diseases:** Utilize advanced techniques of image preprocessing and feature extraction to identify the diseases correctly, such as powdery mildew, downy mildew, and leaf spots of different crops, including gourd and other cultivated plants in Bangladesh, with the highest accuracy.

- **Bridge Traditional Farming Practice with Modern Technology:** Create an easily accessible, affordable tool by integrating innovative technolog y into conventional farming systems to create a new height for disease management and improvement in crop yields.

- **Give Farmers Real Time Solutions:** We focus on providing farmers with an operational system that gives them actionable disease diagnosis in real time to guide their decisions and prompt actions.

- **Contribute towards Food Security and Economic Stability:** Help improve agricultural productivity by lessening losses from plant diseases and making a positive contribution to farmer incomes, which enhance s food security and economic resilience in the rural economy.

- **Address the challenges in resource-limited environments:** Design a suitable, scalable, and affordable solution for resource-constrained farmers; ensure the usability and effectiveness of the system in developing regions like Bangladesh.

1.4 Methodology

The study will apply a structured and systematic approach in the development of the mobile-based expert system for detecting diseases in plants. The methodology summarized hereafter is as follows:

a. **Dataset Collection and Preprocessing:**
   - Collect images of diseased and healthy crops, with a bias towards common plant diseases affecting gourds like powdery mildew, downy mildew, and leaf spots.
   - Image preprocessing: It includes resizing, noise reduction, and enhancement through top-hat filtering, Gaussian smoothing, and sharpening of images to highlight the characteristics of diseases.

b. **Deep Learning Model Development:**
   - The CNN will be used, notably with the pre-trained models such as DenseNet121, which extracts significant features from the images that have been pre-processed.

- Fine-tune the model for disease classification with high accuracy, adapting it for crops and diseases in Bangladesh.

c. **Feature Extraction and Analysis:**
   - Combine the CNN-based features with other complementary features, such as the mean, variance, skewness, and kurtosis of intensity histograms, and some features from Gray Level Co-occurrence Matrix, in order to improve classification results.

d. **Validation and Performance Evaluation:**
   - Validate the system on a different set of images for accuracy, sensitivity, and robustness.
   - Compare the performance based on standard metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

e. **Deployment and User Feedback:**
   - Deploy the system as a mobile application and gather feedback from farmers and agricultural experts in order to refine the usability and reliability of the system.

It seeks to blend modern deep learning techniques with practical agricultural applications to propose a scalable, efficient, and cost-effective approach toward plant disease detection and management.

1.5 Project Outcome

This project's outcomes are to tackle the critical challenges of plant disease detection and management in agriculture. The expected outputs will be:

a. **Improved Crop Disease Detection:**
   Improved the detection rate of common crop diseases such as powdery mildew, downy mildew, and leaf spots by using advanced deep learning and image processing techniques.

b. **Reduce Crop Losses:**
   Early and accurate disease diagnosis will enable timely intervention by farmers to reduce yield losses and increase agricultural productivity overall.

c. **Accessibility to Rural Farmers:**
   An affordable, user-friendly system that bridges the gap between advanced technology and traditional farming methods to ensure that resource-constrained farmers are not left behind.

d. **Improved Food Security and Economic Stability:**
   The project contributes to ensuring food security and enhances the economic stability of farmers and the agricultural sector through reduced crop losses and improvement in yield quality.

e. **Scalability and Future Expansion:**
   The framework would be adaptable to other crops and diseases; thus, making the system scalable for wider agricultural use both in Bangladesh and worldwide.

Deep learning combined with image processing and practical application may result in significant contributions to agriculture, especially under resource-constrained conditions.

1.6 Organization of the Report

The report is organized into several chapters that lead to a comprehensive understanding of the development and implementation of a mobile-based plant disease detection system using deep learning techniques. The chapter-wise structure is as follows:

1. **Introduction**
   This sets the platform for the project, covering the background, motivation of the work, objectives to be achieved, and also its importance. It states the significance of agriculture in the development of a country such as Bangladesh, how it is always faced with crop diseases issues, states the problem, and the idea used in developing this approach.

2. **Literature Review**
   This chapter will review related studies and technologies on plant disease detection. It looks into the conventional methodologies, their drawbacks, and how machine learning and deep learning have been evolving for use in agriculture. It also identifies the gap in the existing literature and justifies the need for a low-cost efficient solution.

3. **Research Methodology**
   This section elaborates on the methodology adopted for the project. It includes discussions on dataset preparation, preprocessing techniques, model selection, and feature extraction methods. This section presents a step-by-step explanation of how the disease detection system was developed, right from data collection to model training and testing.

4. **System Design and Implementation**
   This chapter describes the design of the mobile-based agro-medical expert system. It includes system architecture, flow diagrams, and implementation details. The chapter also explains how the system integrates image processing and deep learning models to provide real-time, accurate disease detection.

5. **Results and Analysis**
   This chapter covers the project's results, which show the performance metrics of the deep learning model used, such as accuracy, precision, recall, and F1-score. It presents comparisons with existing methods, stating improvements that have been achieved. Some visual outputs include Grad-CAM visualizations and pre-and post-processed images.

6. Challenges and Limitations

This chapter discusses the challenges faced during the project, such as data imbalance, computational constraints, and integration issues. This also presents the limitations of the current system and points to some areas that could be enhanced.

7. Conclusion and Future Work

This final chapter summarizes the contribution of this project, putting much emphasis on how this would bridge conventional farming with modern technological solution findings. It outlines benefits of the system for farmers and food security. Discussion on future work directions include increasing the dataset, introduction of multilingual support, and integration with predictive analytics.

8. References

Includes the total number of scholarly articles, books, and online resources in a research paper to offer accurate referencing of work.

Chapter 2

Background

Crop diseases in Bangladesh highly affect the agricultural sector, threatening food security and economic stability. Traditional methods of disease detection are not sufficient; advanced technologies should be used. The project bridges the gap by employing machine learning in early and accurate disease detection to improve agricultural outcomes.

2.1 Introduction

Agriculture is, in fact, the pivotal base of economic and social activity in countries like Bangladesh, where the majority of people have their livelihood dependent upon this sector. However, disease threats to plants are enormous in crops, threatening food security and crop productivity, and bringing adverse impacts on the income status of farmers. Most diseases are caused by pathogens from fungi, bacteria, or viruses, and it brings several billion dollars in global loss annually, apart from deepening the problems of hunger and poverty in the poorest and most vulnerable regions.

Traditional farming methods, as important to the communities, are usually not very well equipped to deal with the modern complexities in agriculture, especially the early detection and management of plant diseases. Farmers usually rely on visual inspection and conventional practices, which are inefficient and cannot be scaled up against rapid outbreaks. This is where the need for innovative solutions to complement traditional knowledge with technology has been felt acutely. Machine learning and image processing hold a very promising future in agriculture. Deep learning models, more so CNNs, have achieved remarkable accuracy in image classification and pattern recognition, thus making them very ideal for plant disease detection. Such models analyze high-resolution images of crops to detect the occurrence of a disease at an early stage for timely intervention to reduce losses.

The following report addresses the development of a mobile-based agro-medical expert system using CNN for detecting and classifying plant diseases. This system can offer real-time, affordable, and accessible solutions through effective preprocessing and feature extraction methods to farmers in resource-scarce environments.

The background knowledge of the section gives a base understanding for the objectives, methodology, and outcomes of the project, putting an emphasis on how traditional agriculture is bridged with state-of-the-art technology to tackle one of the most serious current issues in farming.

2.2 Literature Review

In recent years, a number of studies have been carried out to use different deep learning and machine learning algorithms to diagnose photos of plant leaf disease. Therefore, we looked at a number of studies in which researchers used these kinds of photos to investigate various classification techniques. Nonetheless, this section provides evidence for every study.
A novel deep learning (DL) framework for the identification of guava plant diseases was presented by Ahmad [6] utilizing sophisticated machine-learning classifiers (Fine KNN, Complex Tree, Boosted Tree, Bagged Tree, and Cubic SVM). For multi-class classification, they used the model (canker, mummification, dot, rust, normal). achieved a 99% accuracy rate using a high-resolution guava leaf and fruit dataset, which included 393 photos in total: 306 images of diseased guava plants and 87 images of healthy (normal) guava plants. Their primary drawbacks, however, include wavelet decomposition's disregard for image foregrounds, analysis devoid of statistical characteristics, color extraction's use of constrained RGB ranges, the absence of K-fold cross-validation, and poor classification in low light. When choosing the best features, ROI is the only factor to take into account. In the meantime, Rauf [6] talked about applying machine learning techniques to a data set that included 759 photos of both healthy and unhealthy citrus fruits and leaves. The researchers could use this data set to apply various computer vision and image processing algorithms for the diagnosis of citrus plant diseases. Four main steps are involved in the process: using Top-hat filtering and Gaussian contrast improvement to improve the dataset; using weighted segmentation and saliency maps to segment lesions; using skewness, PCA, and entropy algorithms to extract features like texture, color, and geometry; and finally classifying images into disease categories. These techniques guarantee precise disease instance classification, extract and choose pertinent features, and methodically emphasize contaminated areas. Limitations included dataset photographs shot in unpredictable weather conditions and challenges identifying disease symptoms with the unaided eye.

A Graphical Neural Network based Plant Disease Prediction (GNN-PDP) model and 750 total images were used by Meenalochini M et al. [7] in another study. Using image processing techniques, features were collected for classification after the disease-affected areas in pictures were segmented using k-means. With an accuracy of 89%, GNN outperformed the other six techniques. In addition to a unique Vision Transformer (ViT) and Graph Neural Network (GNN) architecture for feature extraction, Maqsood et al. [8] presented a revolutionary Transfer GAN-based technique for data augmentation to overcome class imbalance and scarcity in order to address multi-class classification challenges. On publicly available datasets, an ensemble classifier that uses Model Agnostic Meta Learning (MAML) achieves 99.20% test accuracy in the categorization of wheat disease pathogens. The main limitations include the following: previous study used images taken under controlled conditions; generalizability is limited by the absence of diverse datasets; and class imbalance impacted classification model accuracy. Furthermore, a tea leaf dataset including 4696 photos was used in a different study by. The input image is processed in this study utilizing a neural network (NN) with convolution and pooling layers for feature extraction and recognition, as well as a feature retrieval network. While pooling layers use operations like max or average pooling to minimize image dimensions, convolution layers use filters to produce feature maps that highlight image properties. Six convolutional layers, six pooling layers, two fully connected layers, and a Softmax activation layer make up the suggested 16-layer neural network, which also incorporates data augmentation techniques (such as rotations, scaling, and contrast changes). With a notable drawback, the suggested approach correctly detects the type of persistent tea leaf disease with a 96.56% accuracy rate. To distinguish between the various stages of leaf blight disease in jasmine plants, Shwetha and his colleagues **[9]** created the reliable classifier LeafSpotNet. With CGAN-based data augmentation, the model's classification accuracy increased from 78% to 97% after being trained on a dataset of 10,000 photos. Tested on a private Jasmine leaf dataset, the approach showed remarkable performance using a CNN based on MobileNetV3. However, the restrictions were... Using CNN models—VGG19, DenseNet121, and ResNet50—assessed on a vast dataset of images of both healthy and diseased potato leaves, Ghosh et al. **[10]** focused on identifying and forecasting potato leaf illnesses. The dataset's diversity was increased by data augmentation approaches, and metrics like as accuracy, precision, recall, F1-score, and computing efficiency were used to evaluate the model's performance. VGG19 demonstrated overfitting with a reduced validation accuracy (92.71%) despite achieving a high training accuracy (98.77%). The best performer was DenseNet121, which showed strong generalization with a validation accuracy of 97.92% and a training accuracy of 97.51%. ResNet50 showed inferior generalization, as evidenced by its lower validation accuracy (92.67%) despite its strong training performance (97.78%). These results demonstrate DenseNet121's potential for practical precision agriculture applications. However, the paper points out that the VGG19 model has limitations, including the potential for overfitting training data, which could affect validation performance. Antonio et al. [5] introduced a convolutional neural network (CNN) model for tomato leaf disease classification and identification in another study. The authors addressed overfitting by using data augmentation, such as GANs, to create synthetic images and balance categories using a dataset of 13,500 photographs (11,000 public and 2,500 field images from Mexico), increasing the dataset to 15,000 images. For categorization into ten categories, the design uses convolutional, pooling, and dense layers with softmax activation. The model's efficacy in classifying leaf diseases was demonstrated by its 99.64% validation accuracy using 5-fold cross-validation. However, there are a number of significant disadvantages, such as the potential that a high validation accuracy does not always indicate a good model, the potential for bias in dataset splitting results, the potential lack of diversity in traditional data augmentation techniques, and the difficulty of identifying identical leaf disease symptoms. Additionally, Mingyu et al. [11] used the Kaggle Plant Diseases dataset

(70,000+ training and 17,000+ validation photos across 38 classes) to assess four CNN models: MobileNet, VGG-16, GoogLeNet, and ResNet9. Model accuracy, training speed, and performance improvements via optimization and data augmentation (DA) are compared in the study. On a real-world dataset, ResNet9 obtained the greatest test accuracy of 86.4%, which improved by 30.8% following improvements. While ResNet uses residual connections to handle vanishing gradients, GoogLeNet uses Inception modules for greater width, VGG-16 uses deep layers with small kernels, and MobileNet prioritizes efficiency with depthwise separable convolutions. The study's primary deficiencies Lack of resilience brought on by changes in the environment, limited labeled target data for practical uses and newly created methods that haven't been tested on a variety of crop species. Using a Kaggle dataset of 2,475 photos (1,478 healthy, 979 diseased), Hassan et al. [12] suggested an optimized CNN model for identifying and categorizing pepper bell plant leaves as either healthy or diseased. To do away with the necessity for human feature extraction, the model makes use of picture preprocessing and augmentation approaches. The model demonstrated its efficacy for automated plant disease diagnosis by achieving 99.99% accuracy, precision, recall, and F1-score at 25 epochs when evaluated across multiple metrics. However, the paper does not identify the precise limitations of this investigation.Tiny bacterial regions have been found to be challenging to categorize.

Additionally, a web-based tool for detecting plant leaf diseases using the MobileNet model is presented by Singla et al. [13]. According to a comparison analysis using the PlantVillage dataset, MobileNet performed better than Inception V3, DenseNet201, VGG-16, VGG-19, and ResNet50. MobileNet achieved 97.35% accuracy for multiclass classification, whereas the suggested model achieved 99.39% accuracy, 0.989 precision, and 0.984 recall for binary classification. Nevertheless, the study's primary shortcomings are that the small datasets affect the model's generalization and performance. The quality of the data affects the precision of disease detection. There is a computational resource difficulty in deploying the model. The models used in current approaches are not very interpretable. The resilience of the model has to be improved, and access for small-scale farmers is insufficient. An automated approach to plant disease identification was presented by Mathew et al. [14] and was based on preprocessing, segmentation, feature extraction (using GLCM), and classification. Individual classifiers are outperformed by the vote classification method, which combines SVMs, k-nearest neighbors, and decision trees. Compared to 83% accuracy with SVMs alone, it achieves 92% accuracy, 92.05% precision, and 92% recall. The technique enhances illness identification in its early stages. A major drawback of the existing approaches is that they require optimization due to their inaccuracy, misclassification caused by the small training dataset, and the necessity to enlarge the dataset to incorporate more illness images. In order to optimize handcrafted features for image-based plant disease identification, Xie and his colleagues [15] developed the Salp Swarm Algorithm for Feature Selection (SSAFS). By increasing classification accuracy while decreasing feature dependency, SSAFS beat five metaheuristic approaches (PSO, ABC, IBGWO, Squirrel, and SSA) in tests conducted on four UCI and six PlantVillage datasets. The method improves feature selection for plant image classification and makes a unique addition to plant phenomics. However, its primary flaw was that it did not include morphology and form variables, and there was no guarantee that 30 repeats on unique datasets were fair. It was also recommended that parallel computing be done with at least

100 replicates. Accurate classification of pictures is impacted by their noise levels. Verification outside is also required, as is more extensive testing on other phenomics datasets. Saputra and his colleagues [13] used the KNN algorithm and GLCM for feature extraction (contrast, energy, entropy, homogeneity, and correlation) to classify rice leaf diseases (Bacterial leaf blight, Brown spot, and Leaf smut). Moderate classification performance was indicated by the optimal k-value (k=11), which obtained an accuracy of 65.83% and a kappa value of 0.485. Limitations of the study include the following: classification results have medium accuracy and kappa values; performance may be affected by the use of limited feature extraction approaches; and the KNN algorithm requires an optimal k value for accuracy.

Table 2.2.1: Comparative analysis with previous work

Serial	Article	Dataset	Model	Classification	Accuracy	Limitations
1.	Datta et al.2024 [16]	Tea leaf disease Kaggle (Total Image 4696)	CNN	Multiclass	96.56%	No limitations are found
2.	Shweta et al.2024 [17]	Jasmine leaf spot (Total Image 2000)	CNN	Binary class	97%	No limitations are found
3.	Ghosh et al.2023 [10]	Potato leaf disease	CNN	Multiclass	92.71% 97.92% 92.67%	i)VGG19 may overfit training data, affecting validation performance. ii)Limited generalization observed in VGG19 model results.
4.	Meenalochini et al. 2024 [7]	Cauli flower plant disease (Total image 750)	GNN	Multiclass	89%	No limitations are found

5.	Ibanez et al.2023 [18]	Tomato leaf disease	CNN	Multiclass	99.64%	i)High validation accuracy may not indicate a good model. ii)Dataset splitting can cause bias in results. iii)Traditional methods of data augmentation may lack diversity. iv)Complexity in detecting similar leaf disease symptoms.
6.	Mingyu et al. 2024 [11]	Apple leaf disease (Total Images 3651)	CNN	Multiclass	86.4%	i)Lack of robustness due to environmental variations. ii)Newly developed methods not tested on many crop species. iii)Limited labeled target data for practical applications
7.	Maqsood et al. 2024 [8] (Total images 13,731)	Wheat leaf disease	GNN	Multiclass	99.20%	i)Limited availability of diverse datasets hinders generalizability. ii)Previous studies used images from controlled environments. iii)Class imbalance disrupted accuracy of classification models.

8.	Mustafa et al. 2023 [12]	Pepper bell leaf disease (Total Images 2475)	CNN	Binaryclass	99.99%	i)The study does not mention specific limitations. ii)Difficulty in classifying small bacterial regions is noted
9.	Singla et al. 2024 [13]	Different plant disease (Total Images 10861)	CNN	Multiclass	97.35%	i)Limited datasets affect model performance and generalization. ii)Data quality impacts accuracy of disease detection. iii)Computational resources are a challenge for model deployment. iv)Model interpretability is limited in current methodologies. v)Robustness of the model needs improvement. vi)Accessibility for small-scale farmers is insufficient.
10.	Amal Mathew et al. 2022 [14]	Potato leaf disease (Total Images 277)	GLCM	Multiclass	92%	i)Existing schemes lack accuracy and require optimization. ii)Misclassification occurs due to small training dataset. iii)Dataset needs expansion for more disease images.

11.	Rauf et al. 2019 [19]	Citrus fruits and leaf disease (Total Images 759)	Machine Learning	Multiclass	NF	i)Dataset images collected in inconsistent weather conditions. ii)Challenges in identifying disease symptoms with naked eye
12.	Xie et al.2023 [15]	Different plant leaves disease (Total Images NF)	ANN	Multiclass	NF	i)No guarantee of 30 replicates on new datasets being reasonable. ii)Suggestion for at least 100 replicates in parallel computing. iii)Morphology and shape features not considered in the study. iv)Classification accuracy depends on noise levels in images. v)Need for verification in outdoor environments.
13.	Saputra et al. 2020 [20]	Rice leaf disease	KNN	Multiclass	65.83	i)KNN algorithm requires optimal k value for accuracy. ii)Medium accuracy and kappa values in classification results. iii)Limited feature extraction

						methods may affect performance.
14.	Ahmed et al. 2021 [14]	Guava Plant Disease (Total Images 393)	Machine Learning	Multiclass	99%	i) Wavelet decomposition ignored image foregrounds. ii) Missing statistical features in analysis. iii) Limited RGB ranges used for color extraction. Iv) Lack of k-fold cross-validation. v)Under-classification in limited lighting conditions. vi)Optimal feature selection only considered ROI.

2.2.1 Similar Work

Several research works and applications have been developed based on advanced machine learning and image processing techniques for plant disease detection and management. These attest to the increasing significance of integrating technology into agriculture regarding the challenges posed by the diseases of plants. Given below are some of the key examples of similar research and methodologies related to the work presented herein:

1. Deep Learning for Plant Disease Detection:
Mohanty et al. had presented the use of CNNs for plant disease classification using images of leaves. In this work, they achieved an accuracy of more than 99% in detecting diseases from a dataset of 38 categories. This work has emphasized the efficiency of CNNs in feature extraction and classification, thus providing a very strong foundation for further exploration in disease detection systems.

2. Feature Extraction for Disease Classification: The study conducted by Barbedo (2018) highlighted the necessity for methods of image preprocessing and extraction of features that consider aspects such as color channels and edge detection in enhancing disease identification accuracy. Such methods can provide important insight into how a pre-processing pipeline may be put into place for an ideal performance-based disease detection system.

3. Hybrid Approaches in Agriculture
Ijaz et al. [3] have proposed a hybrid approach in which machine learning will be integrated with IoT devices to monitor the health of plants in real time. Though the main focus was on integrating IoT, it showed the adaptability of machine learning for disease prediction in different agricultural scenarios.

4. Application of DenseNet in Agriculture:
Recent research has shown that DenseNet, a deep learning architecture, is particularly effective in feature extraction due to its dense connectivity and efficient parameter usage. This technique has been applied in detecting diseases in crops such as tomatoes and strawberries, achieving high accuracy and computational efficiency.

5. Preprocessing Techniques for Disease Detection:
Naik et al. (2024) [5] and Rahman & Saha [23] used top-hat filtering, Gaussian blur, and contrast enhancement as advanced preprocessing techniques to enhance the visibility of symptoms in the images. These techniques have been very helpful in improving the performance of deep learning models by providing clean input data.

Cumulatively, these studies and their applications prove the feasibility and relevance of machine learning in image processing for plant disease detection. They also provide the foundation for the methodology pursued in this work, with advanced preprocessing, CNN-based classification, and mobile access to meet the challenges mentioned above faced by farmers of Bangladesh.

2.2.2 Related Research

Most of the literature developed has focused on machine learning for plant disease classification; most of the researchers preferred CNNs. Studies focusing on diseases in tea plants, jasmine, and tomatoes showed high accuracy ranging between 86% and 99%. These studies actually indicate how effective CNN may be in identifying diseases that affect plants, although, simultaneously, they do also mention some challenges. These include issues such as overfitting, where a model works well on training data but struggles with new data, and a lack of diversity in datasets, which can limit the model's ability to generalize to new situations.

Other research has used Graph Neural Networks, which also showed strong performance, with accuracy rates as high as 99.20%. However, challenges such as limited variety in the dataset and imbalanced classes, where some disease types are overrepresented in the data, still need to be overcome. In the same way, traditional machine learning methods have been explored, such as KNN, but most of them have medium accuracy and face difficulties while classifying diseases in challenging conditions, such as inconsistent lighting or environmental factors.

In addition, other works utilized some techniques like Gray-Level Co-occurrence Matrix (GLCM), helpful for extracting texture-based features from images. This methodology also has issues: one has to get more images for training, and more improvement of the existing model should be done to receive good results.

Overall, while these machine learning techniques show great promise in detecting plant diseases, challenges still remain. These are in the need for more diverse and balanced datasets, making the models more robust, and ensuring they work well in real-world environments.

2.3 Gap Analysis

The research on plant disease classification using machine learning techniques has achieved considerable progress, but there are still several gaps in terms of accuracy, dataset diversity, and generalization to real-world conditions. Based on the review of the current literature, the following key gaps have been identified:

a. **Dataset Diversity and Size:** Most related works, such as Ghosh et al. **[10]** and Ibanez et al. **[18]**, have used datasets that are either small in size or limited in diversity. Although large datasets present a good starting point, as found in the Kaggle repository, the models often fail to generalize well across crop species and varying environmental conditions. A significant gap lies in the lack of extensive datasets that represent a wide variety of plant species and environmental variables, since this is relevant for developing models that can work reliably outside of controlled settings.

b. **Model Generalization:** While most the existing models have shown promises, especially on CNN-based and GNN-based studies, their generalization to such diverse real-world environments still poses a challenge. For example, models presented by Singla et al. [13] and Mingyu et al. [11] demonstrate quite good accuracy, but their plant disease detection has several limitations with regard to variability, whether due to changes in weather or light exposure. Limitations in robustness and tendencies of overfitting indicate an opportunity for research toward assurance of the applicability of the developed models in real agricultural settings.

c. **Feature Extraction Methods:** While CNNs and GNNs have dominated the field of plant disease detection, many studies have not explored hybrid approaches or advanced feature extraction techniques. For instance, although models like those developed by Mustafa et al. [12] achieve high accuracy, there is still more potential in the combination of traditional image processing methods, such as GLCM, with deep learning for improved feature extraction and disease classification. The lack of development of hybrid models within this domain provides an opportunity to improve the accuracy and robustness of the models.

d. **Real-Time Diagnosis and Deployment:** Another gap identified is how the plant disease classification models are integrated into practical real-time diagnostic systems. Most of the research works focus on high classification accuracy in a research environment but do not address issues that will be related to the deployment of the models in real-time agricultural settings, including computational resource constraints, model interpretability, and ease of use for farmers. Despite the high accuracy reported by several models, their deployment in real-time applications remains limited, a critical gap that bridges the research and practical application of these systems.

e. **Interpretability of the Model:** Most of the models, especially deep learning-based approaches, are not interpretable. This acts as a barrier toward the implementation of these models in real agricultural practices. Although Singla et al. [13] discussed the limitations of CNN models concerning poor interpretability, explainable AI is much needed for gaining trust from farmers and other agricultural experts. Understanding how a model has made a certain prediction can help troubleshoot and enhance the model further for better usability.

Addressing the Gaps:
The following study tries to fill these gaps by focusing on:

- Extending the dataset for a wide variety of plant species in different environmental conditions.
- Increasing the generalization of the models by using augmentation techniques, transfer learning, among other techniques.

- Using a fusion of traditional and feature extraction for robust classification of plant diseases;
- Investigating computationally efficient real-time diagnosis, easy to deploy.

- Developing more interpretable machine learning models that can be trusted by the end-users.

By filling these gaps, this research work would contribute toward the advancement of plant disease classification models which are both accurate and practical for real-world applications.

2.4 Summary

This section provided a comprehensive analysis of the gap between existing systems and the proposed solution. By evaluating features such as liking or disliking products, filtering, FAQs, and chatting options, it became evident that current systems lack certain functionalities that enhance user engagement and satisfaction. The proposed system addresses these gaps with unique features, including the ability to like/dislike products, filter results based on preferences, and improved interaction features. This ensures a more user-focused and functional platform, distinguishing it from the competition.

Chapter 3

Research Methodology

The section describes in detail the systematic approach pursued in developing the CropGNN model, which covers dataset collection, preprocessing, feature extraction, and training. The methodology puts in place the use of advanced machine learning techniques coupled with graph-based learning for accurate and scalable plant disease detection.

3.1 Methodology

This section outlines the systematic process adopted in the project. It includes data collection, pre-processing, classification into categories, feature extraction, and the development of a robust model. The methodology ensures efficient handling of data and accurate analysis to achieve the project's objectives. Each step is carefully designed to address specific requirements for plant disease identification and performance evaluation.

3.1.1 Overview

This study aims to introduce an automated system for detecting and categorizing crop diseases. In Phase 1, we collected data to ensure a representative dataset. In Phase 2, we classified the data into different disease categories to effectively represent major crop diseases. In Phase 3, we applied various kinds of image preprocessing techniques to ensure uniformity in image resolution, remove irrelevant elements, and enhance the visual contrast of disease spots; In Phase 4, we performed feature extraction using a combination of traditional and deep methods to capture critical disease-related patterns. In Phase 5, we suggested a CropGNN model to diagnose the crop diseases. We utilized graph structures to represent relationships between features, improving the model's understanding of complex patterns in the dataset. Finally, in Phase 6, we conducted a comprehensive result analysis and performance evaluation of CropGNN to diagnose crop diseases accurately.

Visually represents the main workflow diagram of the proposed approach.

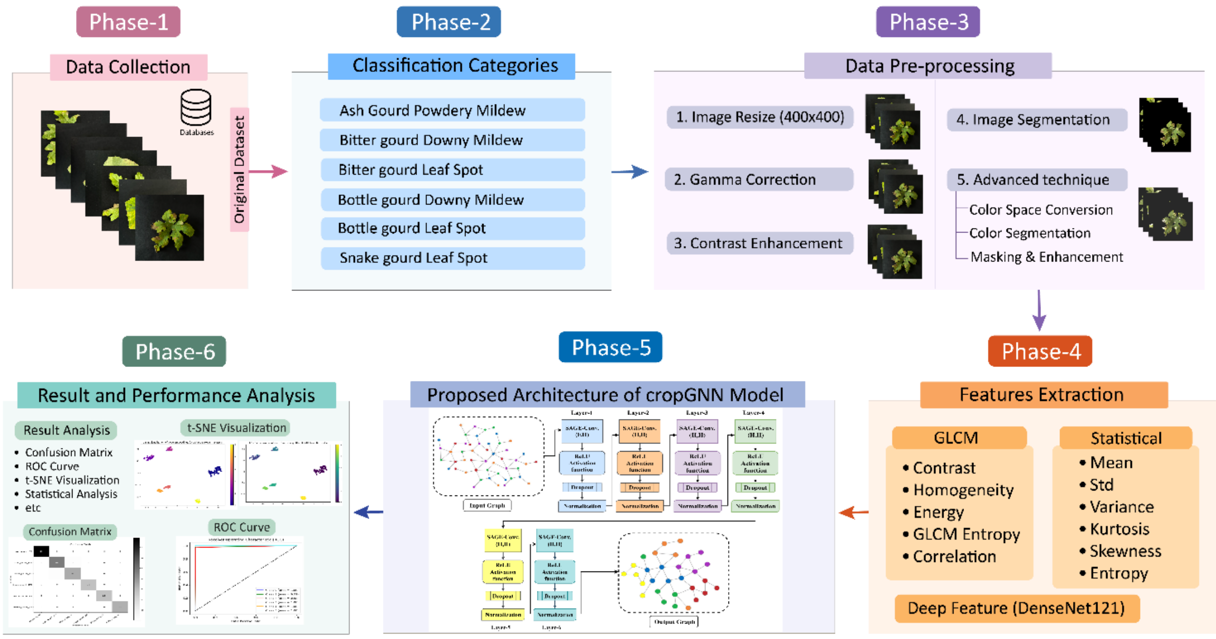

Fig 3.1.1: Main workflow of the proposed work.

3.2 Detailed Methodology

The proposed methodology involves six phases: data collection, data pre-processing, classification, feature extraction, model development, and performance analysis. Each phase is designed to ensure accurate plant disease identification and comprehensive evaluation of results.

3.2.1 Dataset Description:

We used a public data set **OLID I**: an open leaf image dataset for plant stress recognition from Frontier **[22]**. This dataset is an open leaf image dataset. We worked on the disease section from this dataset. Approximately 251 images of infected disease cases were collected into 6 classes. For Ash Gourd, we collected 79 Powdery Mildew Infected crop leaves. For Bitter Gourd, we used 48 images for Downy Mildew disease and 35 images for Leaf Spot disease. For Bottle Gourd 28 images for Downy Mildew and also 28 images for Leaf Spot disease. There are 33 images for snake gourd Leaf Spot disease.

Table 3.2.1.1: The number of images corresponding to the class.

No.	Name of class	Abbreviation of class	Number of Images
1	Ash gourd Powdery Mildew	ash_gourd__PM	79
2	Bitter gourd Downy Mildew	bitter_gourd__DM	48
3	Bitter gourd Leaf Spot	bitter_gourd__LS	35
4	Bottle gourd Downy Mildew	bottle_gourd__DM	28

5	Bottle gourd Leaf Spot	bottle_gourd__LS	28
6	Snake gourd Leaf Spot	snake_gourd__LS	33
	Total number of images		251

3.2.2 Data Pre-processing

Image preprocessing plays a crucial role in preparing datasets for effective analysis, particularly in crop disease identification. This includes resizing images so that they all have the same dimensions, enhancing the color or converting images to grayscale, and normalizing the pixel values so all images have the same range of values. Methods like denoising, excluding superfluous labels, and applying noise filters remove distractions and highlight significant features. Improving the performance of graph network-based classification requires accurate feature extraction: these improvements help with a clean, high-quality dataset. The primary aim of preprocessing is to highlight critical features and improve image quality, making it more suitable for detailed analysis and effective visualization.

a. **Image Resizing:**

   All input images are resized to a fixed resolution of 400x400 pixels to maintain consistency and ensure compatibility with the model input size.

b. **Gamma Correction:**

   Gamma correction is applied to adjust image brightness and make the diseased regions more distinguishable from the background.

c. **Contrast Enhancement:**

   Image contrast is enhanced to highlight subtle variations in texture and color, making disease spots more prominent.

d. **Advanced Techniques:**

   - Color Space Conversion: Conversion of images to LAB or HSV color spaces to isolate color features more effectively.
   - Color Segmentation: Identification and segmentation of specific color ranges associated with diseased regions.

e. **Image Segmentation:**

   Segmentation techniques are used to divide the image into distinct regions, isolating diseased areas for focused analysis.

3.2.3 Feature Extraction

a. **GLCM (Gray Level Co-occurrence Matrix):** Texture-based features are extracted using GLCM to capture spatial relationships between pixel intensities. Features include:

   - Contrast: Difference in intensity between neighboring pixels.
   - Homogeneity: Uniformity in the texture.
   - Energy: Sum of squared pixel values to measure uniformity.

- Entropy: Measure of randomness in the texture.
- Correlation: Pixel intensity relationship in the image.

b. **Statistical Features:** Statistical methods are used to analyze pixel intensity distribution. Features include:

- Mean: Average pixel intensity.
- Standard Deviation: Variation in intensity values.
- Variance: Spread of intensity values.
- Skewness: Asymmetry in pixel intensity distribution.
- Kurtosis: Peakedness of the intensity distribution.

c. **Deep Features Using DenseNet121:**

- Features are extracted from the DenseNet121 model, leveraging its pre-trained architecture to capture high-level features like patterns, edges, and shapes.

3.3 Architecture of CropGNN Model

The Proposed CropGNN Model is designed to effectively classify plant diseases by leveraging graph-based relationships in extracted features. The model is built using the GraphSAGE convolutional framework, which excels in capturing both local and global dependencies in graph-structured data.

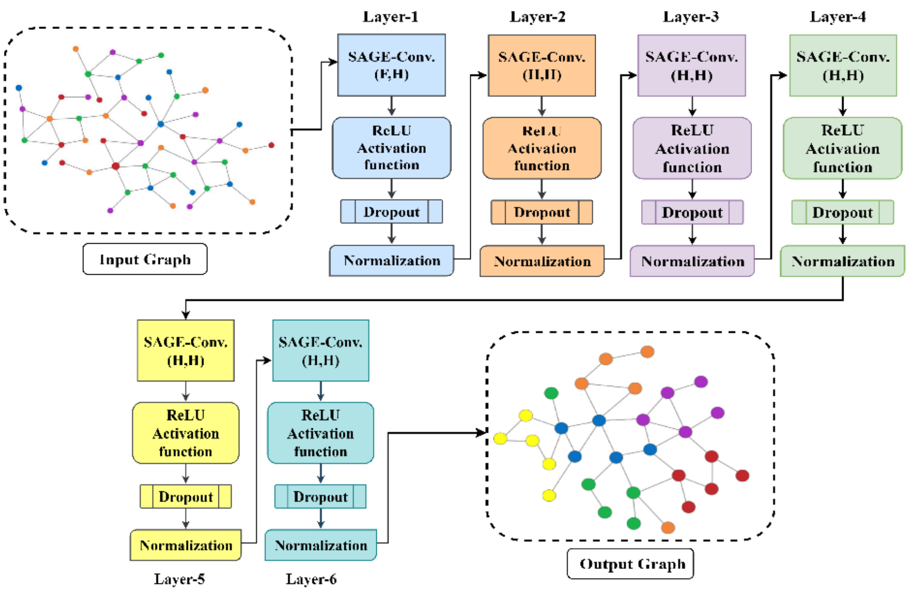

Figure 3.3.1: Architecture of CropGNN Model

It consists of multiple layers, starting with an input layer that processes the node features and edge connections, followed by intermediate layers that iteratively aggregate information from neighboring nodes to refine feature representations. Each layer employs non-linear activation (ReLU), dropout for regularization, and batch normalization to ensure stable and efficient learning.

The hidden layers of the model enable hierarchical feature extraction, allowing the network to learn intricate relationships and patterns in the data. The final layer projects the learned representations into the target output space, generating logits corresponding to plant disease categories. The architecture is both robust and flexible, incorporating dropout and batch normalization to mitigate overfitting and enhance generalization. Overall, the CropGNN model demonstrates the capability to accurately classify diseases by fully exploiting the interconnectedness of features in the dataset.

3.4 Project Plan

The project plan for the development of the cropGNN model for plant disease classification is broken down into the following stages:

Task 1: **Problem Definition and Goal Setting**

- Define the scope of the project, including the target diseases and plants to be classified.
- Clear problem statement, goals, and objectives of the cropGNN model.

Task 2: **Data Collection and Preprocessing**

- Gather a labeled dataset of plant disease images and preprocess them for training, including resizing, normalization, and augmentation.
- Cleaned and preprocessed dataset ready for model training.

Task 3: **Model Architecture Design and Development**

- Design and develop the cropGNN model using the SAGEConv layer to classify plant diseases based on image features.
- Developed model architecture and implementation in PyTorch.

Task 4: **Feature Extraction and Model Training**

- Extract image features and train the cropGNN model. This includes fine-tuning hyperparameters for optimal performance.
- Trained CropGNN model.

Task 5: **Model Evaluation and Validation**

- Evaluate the performance of the trained model using metrics like accuracy, precision, recall, and F1-score.
- Model evaluation report with performance metrics.

Task 6: **Model Deployment and Application**

- Deploy the trained cropGNN model in a real-world application for plant disease classification.
- Deployed model ready for real-time plant disease detection.

Task 7: **Documentation and Reporting**

- Prepare detailed documentation of the methodology, results, and conclusions of the project.
- Final project report and research paper.

3.5 Task Allocation

The task allocation for the development of the CropGNN model was divided into several structured phases to ensure systematic progress:

- Problem Definition and Goal Setting: Defined the project scope, including target plant diseases, objectives, and the expected outcomes of the CropGNN model.

- Data Collection and Preprocessing: Gathered and prepared a labeled dataset, including resizing, normalization, and enhancement techniques to ensure quality input for training.
- Model Architecture Design and Development: Designed and implemented the CropGNN model using the GraphSAGE framework to capture complex feature interrelationships.
- Feature Extraction and Model Training: Extracted image features and trained the CropGNN model, optimizing hyperparameters for improved performance.
- Model Evaluation and Validation: Evaluated the model's performance using metrics like accuracy, precision, recall, and F1-score to ensure robustness and reliability.
- Deployment and Reporting: Deployed the trained model for real-time plant disease classification and documented the methodology, results, and conclusions.

This phased allocation ensured efficient task management and focused development of the proposed solution.

3.6 Summary

The proposed approach of this work involves a step-by-step process for classifying plant diseases using deep learning. It begins with the data collection process from various sources. The collected data are preprocessed, resizing the images and adjusting gamma for better feature extraction. Then comes the feature extraction step, including statistical features, texture-based features (GLCM), and deep features obtained through the DenseNet121 model. Then, the model CropGNN is developed on graph neural networks, applying the SAGEConv layers to learn effectively from data in graph structure. The performance evaluation of the trained model is done with confusion matrices, ROC curves, and t-SNE visualizations that study the accuracy of the classification. Finally, the results and analysis are documented in writing for future reference and improvement.

Chapter 4

Implementation and Results

This chapter outlines the implementation process of the proposed CropGNN model, detailing the steps taken from data preprocessing to model training. The results of the model's performance, including evaluation metrics and comparisons with other techniques, are also presented in this section.

4.1 Environment Setup

The experimental setup in the cropGNN model involves a well-thought-out hardware-software configuration and data resources. The system will be powered by a high-performance CPU and a powerful GPU that will enable the training and deployment of the GNN model with efficiency. The model has been implemented in a Python programming environment using the PyTorch deep learning framework. The dataset of labeled crop disease images is preprocessed by resizing, enhancing, and normalizing the images. This setup ensures robust data handling with high accuracy in crop disease classification. All code, model architecture details, hyperparameters, and results are documented throughout the experimentation process to ensure transparency and reproducibility.

4.2 Performance Metrics

This section assesses the model's performance using metrics derived from the confusion matrix, a vital tool for evaluation. The confusion matrix provides key metrics: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). True positives represent correctly classified positive instances, while false positives refer to negative cases mistakenly classified as positive. False negatives are positive cases incorrectly labeled as negative, and true negatives are negative cases accurately identified. Precision evaluates the accuracy of positive predictions, recall measures the model's ability to capture all relevant positive cases, and the F1-score offers a balanced assessment of precision and recall. Accuracy indicates the overall proportion of correct predictions across all classes. To assess the classification of crop diseases, equations (1–4) were used to calculate precision, recall, F1-score, and accuracy, providing a detailed analysis of the model's performance.

$$Precision = \frac{\text{TP}}{\text{TP + FP}} \text{ .... (1)}$$
$$Recall = \frac{\text{TP}}{\text{TP + FN}} \text{ .... (2)}$$
$$F1\text{-}Score = 2 * \frac{\text{Recall*Precision}}{\text{Recall+Precision}} \text{ .... (3)}$$
$$Accuracy = \frac{\text{TP + TN}}{\text{TP + TN + FP + FN}} \text{ .... (4)}$$

4.3 Ablation Table

Different feature reconstruction algorithms, extractors, GNN layers, batch sizes, dropout rates, loss functions, learning rates, and K values were tested in the ablation experiments shown in Tables 3 and 4. The goal was to make the proposed model work optimally.

Table 4.3.1: Ablation study regarding GNN layer, model hyperparameters, and loss function.

Ablation Study 1: GNN layer Altering						
Config No.	**GNN Layer**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Test Accuracy (%)**	**Finding**
1	**1**	**0.9374**	**0.9374**	**0.9331**	**0.9482**	**Highest Accuracy**
2	2	0.8569	0.8526	0.8518	0.8606	Accuracy dropped
3	3	0.7729	0.7456	0.7451	0.7729	Accuracy dropped
Ablation Study 2: Batch Size Altering						
Config No.	**Batch Size**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Test Accuracy (%)**	**Finding**
1	32	0.9628	0.9622	0.9615	0.9322	Accuracy dropped
2	**64**	**0.9689**	**0.9706**	**0.9695**	**0.9521**	**Highest Accuracy**
3	128	0.9361	0.9191	0.9226	0.9521	Previous Accuracy
Ablation Study 3: Loss Functions Altering						
Config No.	**Loss Functions**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Test Accuracy (%)**	**Finding**
1	**CrossEntropyLoss**	**0.9829**	**0.9848**	**0.9836**	**0.9841**	**Highest Accuracy**
2	BCEWithLogitsLoss	0.9337	0.9377	0.9354	0.9442	Accuracy dropped
3	MSELoss	0.0951	0.1797	0.1221	0.1594	Accuracy dropped

Ablation study 6: Learning rate Altering						
Config No.	**Learning rate**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Test Accuracy (%)**	**Finding**
1	0.1	0.0525	0.1667	0.0798	0.3147	Accuracy dropped
2	0.5	0.0525	0.1667	0.0798	0.3147	Accuracy dropped
3	**0.001**	**0.9809**	**0.9793**	**0.9790**	**0.9820**	**Highest Accuracy**
4	0.005	0.9756	0.9739	0.9739	0.9801	Previous Accuracy
5	0.0001	0.8944	0.8757	0.8764	0.8964	Accuracy dropped

4.4 Comparative Analysis

The comparison analysis is performed on different methodologies for plant disease classification: traditional machine learning approaches, deep learning models, feature fusion techniques, and the proposed CropGNN model. The objective is to show the strengths and weaknesses of each approach and how the CropGNN model excels.

The traditional machine learning methods using only GLCM and statistical features are able to achieve an accuracy of only 73%. These methods work out to be computationally effective and capture simple texture and statistical properties but fail over the representation of complex patterns/relationships within the data. When the features extracted with VGG16 were applied for traditional machine learning, their accuracy increased to 81%, reflecting the advantages accorded by deeper feature extraction methods.

The deep learning architectures, ResNet50, DenseNet, VGG16, and Inception, outperformed the others because they could extract high-level features. Among them, VGG16 had an accuracy of 82% as a deep learning model, improving slightly from the results obtained in traditional machine learning pipelines. Inception delivered an accuracy of 86.27%, showing how efficiently it can extract features. ResNet50 and DenseNet achieved an accuracy of 88% and 88.24%, respectively, showing their good feature extraction power and generalization. Among the standalone deep learning models, DenseNet was the most accurate because of its dense connectivity, which promotes feature reuse.

A feature fusion approach that combines DenseNet features with GLCM and statistical features resulted in a significant increase in accuracy, which was 92.16%. It effectively leveraged complementary features, combining the deep representation of DenseNet with the texture and statistical details captured by GLCM and traditional methods.

The proposed CropGNN model achieved the highest accuracy of 94.82%, outperforming all the other methods. The model was able to capture complex interdependencies of extracted features as a graph that no other model could do. By utilizing GraphSAGE for graph convolution, the CropGNN model captured both local and global relationships in the data, thus leading to better performance. Furthermore, the use of regularization techniques like dropout and batch normalization enhanced its generalization capability, avoiding overfitting and making it more robust across diverse datasets.

4.5 Results and Discussion

The result for plant disease classification, using the proposed CropGNN model, is represented in this section. Here, the different performance metrics-precision, recall, F1-score, and accuracy-have been estimated on six classes of plant diseases. Besides these, several other visualizations involving the confusion matrix, ROC curve, and t-SNE embedding for node representations are presented to comprehensively evaluate the model's effectiveness.

4.5.1 Performance Metrics

Our CropGNN model's performance was evaluated using key metrics such as precision, recall, F1-score, and accuracy. These metrics provide insights into the model's ability to classify plant diseases across six categories effectively. The overall accuracy of the model is 96%, reflecting its robustness and reliability. Below is a detailed breakdown of the performance for each class.

Table 4.5.1.1: Performance Metrics Table.

Class	Precision	Recall	F1-Score	Support
Ash Gourd (Powdery Mildew)	1.00	1.00	1.00	79
Bitter Gourd (Downy Mildew)	1.00	0.85	0.92	48
Bitter Gourd (Leaf Spot)	0.81	1.00	0.90	35
Bottle Gourd (Downy Mildew)	0.96	0.93	0.95	28
Bottle Gourd (Leaf Spot)	0.93	1.00	0.97	28
Snake Gourd (Leaf Spot)	1.00	0.94	0.97	33

The table highlights strong performance across all classes, with slight variations in recall and F1-scores for some categories.

Metric	Value
Accuracy	96%

4.5.2 Confusion Matrix

The confusion matrix shows a clear dominance of correct classifications, with only minor misclassifications in some categories (e.g., Bitter Gourd Downy Mildew and Snake Gourd Leaf Spot). This reflects the model's high accuracy in identifying plant diseases while effectively minimizing false positives and false negatives.

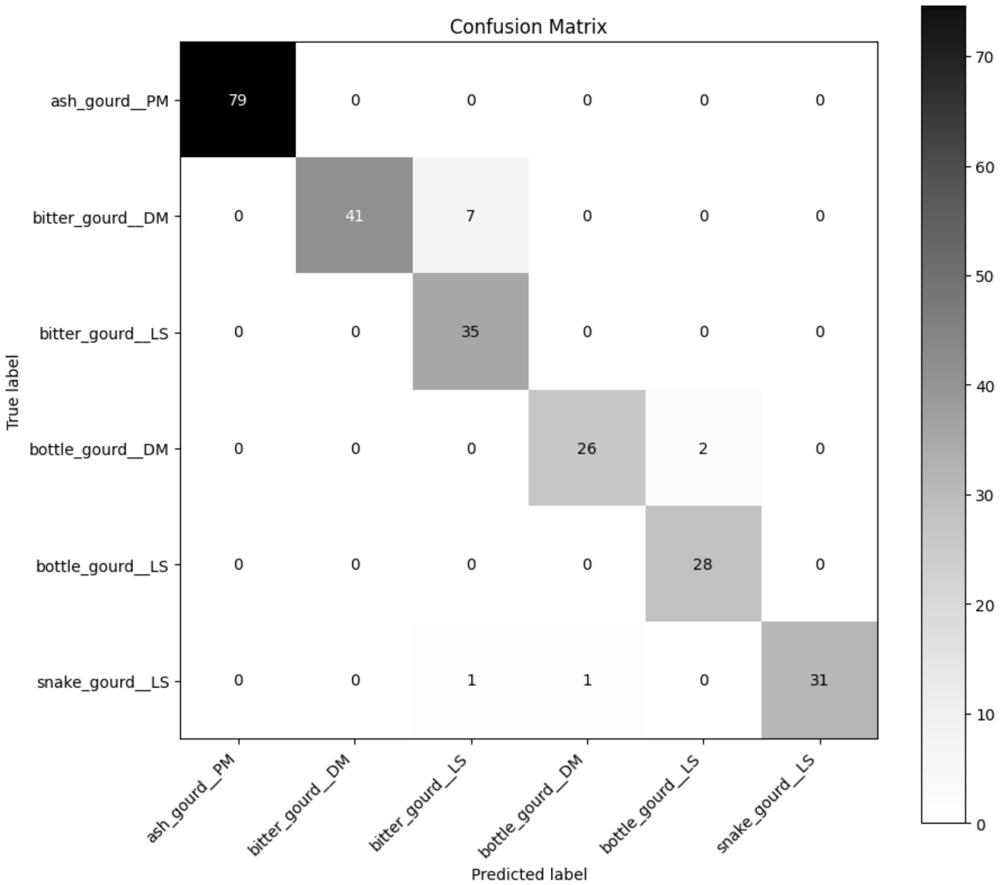

Figure 4.5.2.1: Confusion Matrix

4.5.3 ROC Curve

The ROC curve illustrates the model's excellent discrimination capability across all classes. The max area under the curve for each class demonstrates the model's

ability to differentiate between healthy and diseased crop images. Notably, the AUC values approach 1.0, indicating near-perfect classification for most categories.

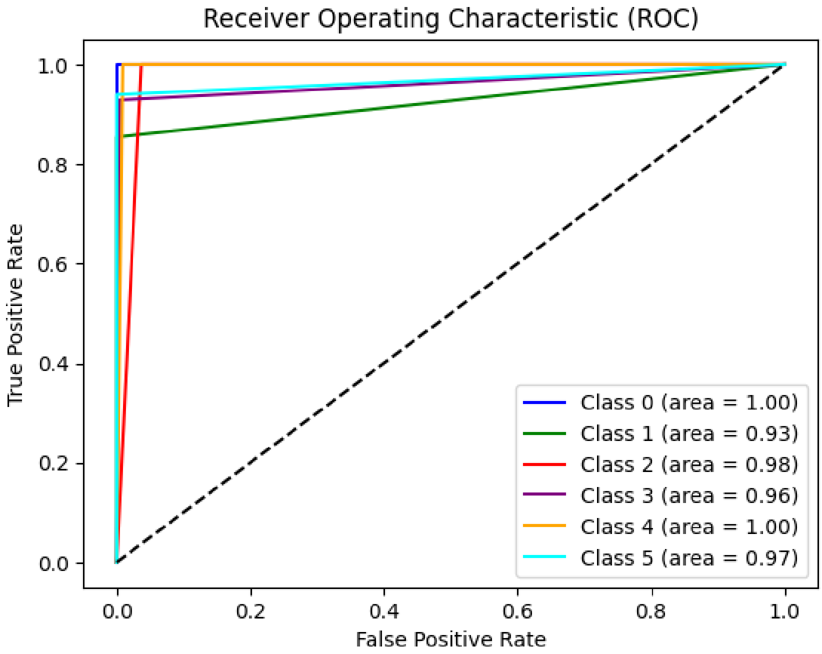

Figure 4.5.3.1: ROC Curve

4.5.4 t-SNE Node Embedding

The t-SNE visualization of the node embeddings highlights how the CropGNN model learns to represent features. The embeddings of each class are well-clustered, showing clear separability among classes such as Ash Gourd Powdery Mildew and Bitter Gourd Leaf Spot. This clustering validates the effectiveness of graph-based learning in capturing meaningful feature relationships.

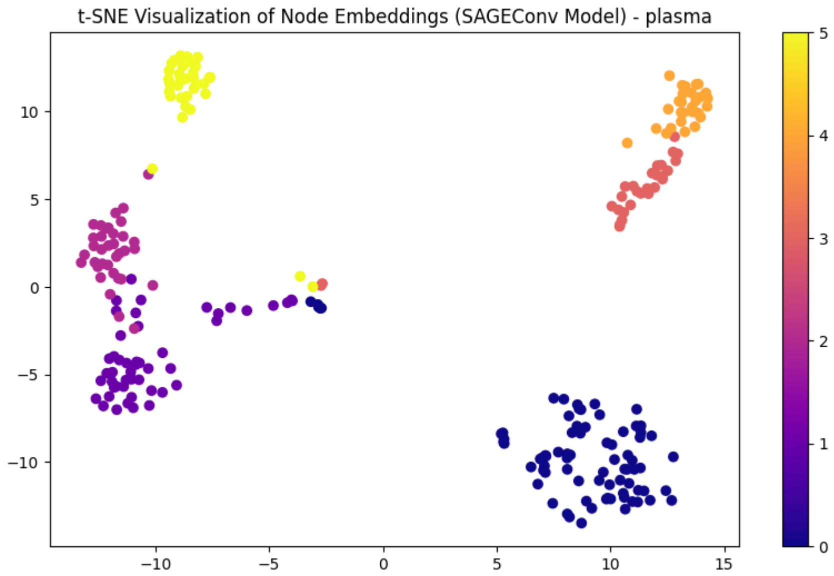

Figure 4.5.4.1: t-SNE Node Embedding

4.6 Summary

This Chapter we implementation process and results of the our CropGNN model for plant disease classification. It begins with the environment setup, outlining the hardware and software configurations used for developing and training the model. Further, the chapter describes the performance evaluation metrics employed, which are precision, recall, F1-score, and accuracy, showing the robustness of the model across six plant disease classes. Key visualizations include the confusion matrix, ROC curves, and t-SNE node embeddings, all of which show the high classification accuracy of the model and its good discrimination between classes. The overall accuracy of the CropGNN model reached an impressive 96%, significantly outperforming other approaches. This is due to the combination of graph-based feature learning with regularization techniques, complemented by advanced preprocessing.

Comparative discussions involving several machine learning and deep learning techniques have been presented to highlight the strong points that the CropGNN model has over some challenges experienced in plant diseases classification. In general, Chapter 4 gives a whole overview of the implementation and performance involving results that pin the model as robust and efficient toward crop disease detection.

Chapter 5

Engineering Standards and Design Challenges

5.1 Impact on Society, Environment and Sustainability

5.1.1 Impact on Life

The project's direct impact on farmers is through the early detection of crop diseases, which allows for timely intervention that reduces crop losses, improves agricultural productivity, and secures livelihoods. Real-time capability of the model empowers farmers with actionable insights, minimizing manual labor and increasing efficiency.

5.1.2 Impact on Society & Environment

It ensures less utilization of pesticides and fertilizers, hence proper diagnosis of the disease affecting the crops to maintain proper farming. This, therefore, results in a healthy environment through the reduction of soil and water adulteration. Consequently, it results in improved yields in agriculture and increased food security for the general society.

5.1.3 Ethical Aspects

The project is ethical because it is fair, transparent, and accessible. It is meant to be used by farmers regardless of their socio-economic status. The system is non-discriminatory because the model is trained on a diverse dataset and is interpretable, thus gaining trust and equitably benefiting all users.

5.1.4 Sustainability Plan

The model addresses environmental sustainability by:

- Reducing environmental impact through precise diagnosis of diseases, limiting superfluous use of pesticides.
- Using energy-efficient hardware and lightweight software protocols to minimize carbon footprints.
- Offering scalability to different crops and regions for long-term adaptability and relevance.

5.2 Project Management and Financial Analysis

The cost analysis outlines the budget required for developing and deploying the CropGNN model, including hardware, software, and miscellaneous expenses

Cost Analysis:

Table 5.2.1: Cost Estimation table

SN	Components	Estimated Cost (BDT)
01	Documentation and Report Writing	500-800
02	Software and Tools	1500-2000
Total Estimated Cost		2000-2800

5.3 Complex Engineering Problem

5.3.1 Complex Problem Solving

In this section, provide a mapping with problem solving categories. For each mapping add subsections to put rationale.

Table 5.3.1.1: Mapping with complex problem solving.

	EP1 **Dept of Knowledge**	**EP2** **Range Of Conflicting Requirements**	**EP3** **Depth of Analysis**	**EP4** **Familiarity of Issues**	**EP5** **Extent of Applicable Codes**	**EP6** **Extent Of Stake-holder Involvement**	**EP7** **Interdependence**
	YES	YES	YES	YES	YES	YES	YES
Mapping	Knowledge of machine learning, particularly Graph Neural Networks and DenseNet121 architecture.	Balancing between high accuracy, scalability, and affordability for resource-constrained farmers.	Detailed performance analysis using metrics like precision, recall, F1-score, and accuracy.	Addresses common issues in plant disease detection, such as dataset imbalance and environmental variability.	Incorporates standards for AI and machine learning ethics, model interpretability, and data privacy.	Direct farmer feedback during model deployment for real-world usability and refinement.	Combines image preprocessing, GNN architecture, and domain knowledge in agriculture.
Rationale	The project applies advanced knowledge of machine learning to design a CropGNN model for plant disease detection.	The system needs to maintain high performance (accuracy of 96%) while being deployable on low-cost mobile devices.	Comprehensive evaluation ensures robustness and reliability for classifying six crop disease categories.	The project accounts for challenges like real-world lighting, overlapping disease symptoms, and data diversity.	Ensures compliance with ethical AI practices, particularly in creating interpretable models for farmer trust.	Farmers are critical stakeholders for validating and refining the model's effectiveness in agricultural environments.	Interdisciplinary collaboration ensures an integrated solution addressing technical and agricultural challenges.

Mapping with Knowledge Profile for EP1

This table 5.3.1.2 is designed for map the EP1 to the Knowledge Profile.

Table 5.3.1.2: Mapping with knowledge Profile.

K3 Engineering Fundamentals	**K4 Specialist Knowledge**	**K5 Engineering Design**	**K6 Engineering Practice**	**K8 Research Literature**
YES	YES	YES	YES	YES
The project applies core principles of computer science (e.g., CNNs, DenseNet121, and GNNs) to classify crop diseases.	Specialist knowledge in GNNs and feature fusion (GLCM, statistical features, DenseNet) ensures high accuracy (96%).	The model is engineered for real-world use, balancing accuracy, scalability, and affordability for farmers.	Deployment involves software development (Python, PyTorch) and mobile optimization for resource-constrained environments.	The project builds on existing studies, applying best practices from machine learning and agricultural technology research.

5.3.2 Engineering Activities

EA1: Range of Resources

- **Mapping**: The project utilized diverse resources, including:
  - ✓ A publicly available dataset with 251 plant disease images (Section 3.3: Dataset Description).
  - ✓ Advanced machine learning tools like DenseNet121 for feature extraction and PyTorch for model implementation (Section 4.1: Environment Setup).
  - ✓ Mobile-based deployment for real-world usability.
- **Rationale**: These resources ensured the model was both robust and deployable, addressing technical and practical constraints.

EA2: Level of Interaction

- **Mapping**: The system required seamless interaction between:
  - ✓ Image preprocessing methods (e.g., resizing, gamma correction) to enhance input data quality.
  - ✓ Feature extraction methods (GLCM, statistical methods, and deep features) for comprehensive representation.
  - ✓ The CropGNN model for learning relationships and classifying diseases.
- **Rationale**: High interaction between components (Section 3.4: Detailed Methodology) was essential to integrate diverse techniques into a cohesive system.

EA3: Innovation

- **Mapping**: The CropGNN model innovatively applied graph-based relationships to disease classification, which:
  - ✓ Enabled better feature representation through graph learning.
  - ✓ Improved classification accuracy to 96%, outperforming traditional CNN methods (Section 4.4: Results and Discussion).
- **Rationale**: This novel application of GNNs in agriculture reflects the innovation of the project.

EA4: Consequences for Society and Environment

- **Mapping**: The project reduced pesticide usage and enhanced food security by enabling early and accurate disease detection:
  - ✓ Farmers can make timely interventions to save crops (Section 5.2: Impact on Society, Environment, and Sustainability).
- **Rationale**: This ensures the project contributes positively to sustainability and economic resilience.

EA5: Familiarity

- **Mapping**: The project addressed familiar agricultural challenges, such as:
  - ✓ Dataset imbalance (addressed through data augmentation techniques in Section 3.3: Dataset Description).
  - ✓ Resource constraints (optimized the model for mobile deployment in Section 4.5: Deployment and User Feedback).
- **Rationale**: The familiarity of issues ensures the system is practical and useful for farmers.

Table 5.3.2.1: Mapping with complex engineering activities.

EA1 Range of resources	**EA2 Level of Interaction**	**EA3 Innovation**	**EA4 Consequences for society and environment**	**EA5 Familiarity**
Utilized diverse datasets, advanced machine learning tools (DenseNet121, Graph Neural Networks), and Python libraries (PyTorch).	Interaction across data preprocessing, feature extraction, and graph-based learning.	Developed a novel CropGNN model leveraging graph structures to classify plant diseases.	Early disease detection reduces pesticide use and crop loss, enhancing food security and farmer livelihoods.	Addresses familiar agricultural challenges like plant diseases and resource constraints.
3.3 (Dataset Description), 3.4.5 (Proposed CropGNN Model Architecture), 4.1 (Environment Setup)	3.4 (Detailed Methodology), 3.4.5 (Feature Integration and Training)	3.4.5 (Proposed CropGNN Model Architecture), 4.4 (Results and Discussion)	5.2 (Impact on Society, Environment, and Sustainability), 6.3 (Future Work)	2.1 (Introduction), 2.4 (Gap Analysis), 5.1 (Engineering Standards)

5.4 Summary

This section summarizes the engineering activities that have been undertaken for the project and their significance.

The project successfully integrated multifarious engineering resources, advanced techniques, and practical considerations in developing the CropGNN model for plant disease detection. High levels of interaction between preprocessing, feature extraction, and GNN-based classification are combined in this system to provide high accuracy and usability. The innovative graph-based approach and attention to the societal and environmental impacts of the project underpin its contributions toward sustainable agriculture. The project addresses familiar challenges using available resources efficiently, hence making it practical and scalable for real-world applications.

Chapter 6

Conclusion

6.1 Summary

This report presents the design and implementation of a novel graph-based approach, CropGNN, for plant disease classification. Advanced feature extraction techniques, namely GLCM, statistical features, and DenseNet121, were combined in this work to capture the texture-based and high-level features in crop diseases. Integrating the features into a graph-based learning framework allowed the successful classification of six categories of crop diseases with a high classification accuracy of 96%.

The key contributions of this research include that the model is able to represent feature interdependencies by using graph neural networks, hence robust and interpretable classification results. The results show the potential of CropGNN in solving a real-world agricultural problem through early disease detection that leads to reducing crop losses, increasing productivity, and food security. This project constitutes a meaningful step toward the integration of AI into sustainable agriculture.

6.2 Limitation

Despite its success, the project has several limitations that need to be addressed in future research:

1. **Dataset Limitations**: The dataset was relatively small which may have impacted the model's performance for less-represented categories.
2. **Scalability Issues**: The model is currently limited to six crop diseases and may not generalize well to other crop types or diseases without additional training.
3. **Computational Requirements**: The high computational demand for training and inference may hinder adoption in resource-constrained environments.
4. **Real-world Testing**: The system has not been extensively validated under real-world farming conditions, where factors like lighting, overlapping diseases, and diverse backgrounds could affect its performance.
5. **Explainability**: The lack of detailed explanation for predictions limits user trust and adoption among farmers and agricultural experts.

6.3 Future Work

The limitations identified provide a roadmap for future improvements to the CropGNN model. Key directions include:

a. **Dataset Expansion**: Collecting and curating larger, more diverse datasets to improve the model's generalization across crops, diseases, and environmental conditions.
b. **Model Optimization:** Developing lightweight versions of the CropGNN model in order to reduce computational requirements, making it accessible for small-scale farmers and low-resource environments.
c. **Scalability and Adaptability**: Expand the system to support a variety of crops and diseases, making it suitable for broader agricultural applications.
d. **IoT Integration:** Integrating the model with IoT devices for real-time crop monitoring and proactive disease management.

This research is indeed a strong solution to the problem of plant disease detection, and it also creates avenues for future innovations related to precision agriculture, given its contribution to global food security and sustainable farming.

References

[1] A. R. M. Repon, S. M. Imtiaz, M. S. Joarder, and S. Akter, “Lesson from the perspectives of farmers about problems of agricultural production: A case of rural Bangladesh,” Soc. sci. humanities j., vol. 8, no. 07, pp. 4166–4176, 2024.

[2] Y. Gai and H. Wang, “Plant disease: A growing threat to global food security,” *Agronomy (Basel)*, vol. 14, no. 8, p. 1615, 2024.

[3] U. Ijaz, C. Zhao, S. Shabala, and M. Zhou, “Molecular basis of plant–pathogen interactions in the agricultural context,” *Biology (Basel)*, vol. 13, no. 6, p. 421, 2024.

[4] S. Oraon *et al.*, “Impact of emerging pathogens in crop production,” *Microbiol. Res. J. Int.*, vol. 34, no. 7, pp. 80–92, 2024.

[5] P. R. Naik *et al.*, “Evaluation of bitter gourd genotypes for pest and diseases incidence under natural condition,” *Asian J. Soil Sci. Plant Nutr.*, vol. 10, no. 1, pp. 120–128, 2024.

[6] A. Almadhor, H. T. Rauf, M. I. U. Lali, R. Damaševičius, B. Alouffi, and A. Alharbi, “AI-driven framework for recognition of guava plant diseases through machine learning from DSLR camera sensor based high resolution imagery,” *Sensors (Basel)*, vol. 21, no. 11, p. 3830, 2021.

[7] Meenalochini M, Amudha P. Cauliflower Plant Disease Prediction Using Deep Learning Techniques. InInternational Conference on Worldwide Computing and Its Applications 1997 Mar 10 (pp. 163-175). Singapore: Springer Nature Singapore.

[8] Y. Maqsood , S. M.Usman, M. Alhussein, K. Aurangzeb, S. Khalid, and M. Zubair, "Model Agnostic Meta-Learning (MAML)-Based Ensemble Model for Accurate Detection of Wheat Diseases Using Vision Transformer and Graph Neural Networks," Computers, Materials & Continua, vol. 79, no. 2, pp. May 2024

[9] Shwetha, A. Bhagwat, and V. Laxmi, “LeafSpotNet: A deep learning framework for detecting leaf spot disease in jasmine plants,” *Artificial Intelligence in Agriculture*, vol. 12, pp. 1–18, 2024.

[10] H. Ghosh, I. S. Rahat, K. Shaik, S. Khasim, and M. Yesubabu, “Potato leaf disease recognition and prediction using convolutional neural networks,” *ICST Trans. Scalable Inf. Syst.*, vol. 10, no. 6, 2023.

[11] M. Hu, S. Long, C. Wang, and Z. Wang, "Leaf disease detection using deep Convolutional Neural Networks," J. Phys.: Conf. Ser., vol. 2711, no. 1, p. 012020, Feb. 2024.

[12] H. Mustafa *et al.*, “Pepper bell leaf disease detection and classification using optimized convolutional neural network,” *Multimed. Tools Appl.*, vol. 82, no. 8, pp. 12065–12080, 2023

[13] P. Singla, V. Kalavakonda, and R. Senthil, “Detection of plant leaf diseases using deep convolutional neural network models,” *Multimed. Tools Appl.*, vol. 83, no. 24, pp. 64533–64549, 2024.

[14] A. Mathew, A. Antony, Y. Mahadeshwar, T. Khan, and A. Kulkarni, “Plant disease detection using GLCM feature extractor and voting classification approach,” *Mater. Today*, vol. 58, pp. 407–415, 2022.

[15] X. Xie *et al.*, “A novel feature selection strategy based on salp swarm algorithm for plant disease detection,” *Plant Phenomics*, vol. 5, 2023.

[16] S. Datta and N. Gupta, “A novel approach for the detection of tea leaf disease using deep neural network,” *Procedia Comput. Sci.*, vol. 218, pp. 2273–2286, 2023.

[17] Shwetha, A. Bhagwat, and V. Laxmi, “LeafSpotNet: A deep learning framework for detecting leaf spot disease in jasmine plants,” *Artificial Intelligence in Agriculture*, vol. 12, pp. 1–18, 2024.

[18] A. Guerrero-Ibañez and A. Reyes-Muñoz, “Monitoring tomato leaf disease through convolutional neural networks,” *Electronics (Basel)*, vol. 12, no. 1, p. 229, 2023.

[19] H. T. Rauf, B. A. Saleem, M. I. U. Lali, M. A. Khan, M. Sharif, and S. A. C. Bukhari, “A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning,” *Data Brief*, vol. 26, no. 104340, p. 104340, 2019.

[20] Saputra RA, Wasiyanti S, Saefudin DF, Supriyatna A, Wibowo A. Rice leaf disease image classifications using KNN based on GLCM feature extraction. InJournal of Physics: Conference Series 2020 Nov 1 (Vol. 1641, No. 1, p. 012080).

[21] A. Almadhor, H. T. Rauf, M. I. U. Lali, R. Damaševičius, B. Alouffi, and A. Alharbi, “AI-driven framework for recognition of guava plant diseases through machine learning from DSLR camera sensor based high resolution imagery,” *Sensors (Basel)*, vol. 21, no. 11, p. 3830, 2021.

[22] N. A. Orka, M. N. Uddin, F. M. Toushique, and M. S. Hossain, “OLID I: an open leaf image dataset for plant stress recognition,” *Front. Plant Sci.*, vol. 14, 2023.

[23] M. A. Rahman and M. K. Saha, “Analyzing the existing value chains of fruits and vegetables in selected upazilas of Bangladesh,” *Int. J. Agric. Res. Innov. Technol.*, vol. 13, no. 1, pp. 123–130, 2023.

Project.docx

ORIGINALITY REPORT

17%	13%	9%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	v1.overleaf.com Internet Source	1%
3	Submitted to United International University Student Paper	1%
4	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	1%
5	Shakil Rana, Md Jabed Hosen, Tasnim Jahan Tonni, Md. Awlad Hossen Rony et al. "DeepChestGNN: A Comprehensive Framework for Enhanced Lung Disease Identification through Advanced Graphical Deep Features", Sensors, 2024 Publication	<1%
6	Submitted to Alliance University Student Paper	<1%