PLANT DISEASE CLASSIFICATION USING DEEP LEARNING APPROACHES

BY

REHNUMA AKTER
ID: 202-1514416

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

Mr. Fourcan Karim Mazumder (FKM)
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Mrs. Sharun Akter Khushbu (SAK)
Senior Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "**Plant Disease Classification Using Deep Learning Approaches**", submitted by **Rehnuma Akter** ID: 202-15-14416 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 **15 July 2024**

BOARD OF EXAMINERS

Dr. Md. Ismail Jabiullah(MIJ) — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Md Assaduzzaman (MA) — **Internal Examiner 1**
Senior Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Ms. Sharun Akter Khushbu (SAK) — **Internal Examiner 2**
Senior Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

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

DECLARATION

I hereby declare that this project has been done by me under the supervision of **Mr.Fourcan Karim Mazumder (FKM), Assistant Professor, Department of Computer Science and Engineering, Daffodil International University**. I 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:

14/07/24

Mr. Fourcan Karim Mazumder (FKM)
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

14.07.24

Ms. Sharun Akter Khushbu (SAK)
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Rehnuma Akter 14.07.24

Rehnuma Akter
ID: 202-15-14416
Department of CSE
Daffodil International University

©Daffodil International University, Bangladesh *iii*

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty for His divine blessing making it possible for me to complete the final year project/internship successfully.

I am grateful and wish my profound indebtedness to **Mr.Fourcan Karim Mazumder (FKM), Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of deep 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 my heartiest gratitude to the **Head of the Department of CSE**, for his kind help in finishing my project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

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

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

ABSTRACT

This thesis extensively evaluates deep learning models for classifying plant diseases and their effects on agriculture, society, and the environment. Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4 are compared for their ability to diagnose and treat agricultural diseases. The models were very accurate, with the EfficientNet B4 model scoring 99.99%. The Modified AlexNet and EfficientNet B0 models followed with 99.6% accuracy. CNN's 99.5% accuracy was impressive. The models also had modest loss values, suggesting good learning and training. With its accuracy, recall, and F1-score measures, EfficientNet B4 reliably classified all disease categories better than any other model. The models' capacity to transform agriculture, enhance crop management, and secure global food security has a major influence on society. The research also examines how deep learning models reduce chemical pesticide consumption and promote sustainable farming. Ethical considerations include data privacy, algorithmic bias, transparency, and responsibility emphasize the need for responsible and fair deployment. A sustainability strategy plans the long-term viability and inclusiveness of agricultural deep learning models. Future research should address dataset limits, examine a variety of data sources, improve model performance, and evaluate ethical issues, according to the thesis. This study highlights the potential of deep learning models in agriculture. It stresses ethical and sustainable methods while using these models to solve social issues and encourage environmental stewardship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

CNN	Convolutional Neural Network
AI	Artificial Intelligence
DL	Deep Learning
F1	F1 Score
EffNet-B0	EfficientNet B0
CPU	Central Processing Unit
EffNet-B4	EfficientNet B4
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
ReLu	Rectified Linear Unit
Adam	Adaptive Moment Estimation
TPU	Tensor Processing Un

CHAPTER 1

Introduction

1.1 Overview

Plant disease Classification are important for maintaining biodiversity and assuring the nutritional quality of human diets. They are staple crops in many nations, particularly in tropical regions, and make a major contribution to calorie consumption [1-3]. With the growing global population, there is a greater need for food, leading to heightened stress on agricultural production. Hover, meeting this criterion is occasionally impeded by unpredictable Lather patterns and challenges related to manual work in the agriculture industry [4]. Historically, technology has been utilized to tackle these challenges, progressing from manual plowing to automated machinery, resulting in enhanced efficiency and output. Categorizing plant diseases, referred to as "Supermarket Botany," is crucial in agriculture to improve agricultural practices and satisfy client demands [5,6]. Identifying and categorizing plant diseases can be challenging because of the similarities in their characteristics, like color, shape, and texture. The manual labor required for tasks like harvesting, sorting, and labeling presents an extra obstacle, hindering the commercialization of plant disease products. In recent years, progress in machine learning and deep learning has enabled the automation of plant disease classification and detection. These technologies can enhance agricultural operations, increase output, and improve the quality of produce [7]. Hover, the complex nature of plant diseases and the multiple classification issue pose significant challenges. In nations like Bangladesh, the implementation of modern, smart agriculture has led to an increase in plant diseases, highlighting the need for automated agricultural solutions. Scientists have proposed novel methods for detecting plant diseases by integrating many analytical techniques to enhance precision. Artificial intelligence and deep learning are powerful methods used to extract important information from images, enabling the development of innovative applications in agriculture [8-10]. Strong datasets, advanced sensor technologies, and efficient feature representation methodologies are still needed to properly address the challenges in plant disease classification. Utilizing new technologies like machine learning and deep learning is crucial to meet the growing food demand and preserve agricultural practices. Automated plant disease classification has the potential to revolutionize agriculture by enhancing efficiency, production, and quality [11]. Continual research and development in this field are essential for addressing existing challenges and maximizing the benefits of technology in agriculture.

1.2 Background and Present State

Advancements in artificial intelligence (AI) have significantly transformed several sectors, such as agriculture and healthcare, in recent years. An example use case is using deep learning models to classify and diagnose plant diseases [12]. This technology provides unparalleled precision and effectiveness in early disease detection, ultimately improving crop management methods and ensuring food security [13]. The agricultural industry has formidable obstacles as a result of the widespread occurrence of plant diseases, which may lead to major global crop losses. Conventional illness detection techniques often depend on visual examination, which is subjective and susceptible to human mistakes [14]. Deep learning models, namely convolutional neural networks (CNNs), have emerged as a potential option for automating and enhancing illness detection procedures [15]. Recent studies have shown that deep learning models, including Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4, are very successful in properly categorising different types of plant diseases. These models use extensive databases of plant photos to acquire knowledge of intricate patterns and can differentiate between healthy and unhealthy plants with great accuracy. Although deep learning models show potential, there are various issues that need to be addressed when using them in agriculture. These include ethical concerns related to the protection of data privacy, the presence of algorithmic prejudice, and the ecological consequences of deploying models. Furthermore, it is essential to guarantee that smallholder farmers and neglected agricultural communities have easy access to and can afford these technologies in order to ensure fair and equal advantages. Current research highlights the profound impact of deep learning on agriculture, which may lead to sustainable agricultural methods, advanced crop management techniques, and greater global food security. This thesis examines and assesses several deep learning structures to contribute to the continuing discussion on improving agricultural technology for the benefit of society. This review provides a foundation for comprehending the importance of deep learning models in tackling crucial issues in agriculture. It also emphasises the research goals and contributions of this thesis in developing the area.

1.3 Problem Statement

The agricultural industry has enduring obstacles concerning the detection and control of plant diseases, which have a substantial effect on crop productivity, food security, and economic stability worldwide. Traditional approaches to disease identification often depend on farmers or agricultural professionals visually examining crops, a process that is subjective, time-consuming, and susceptible to errors. Furthermore, the growing intricacy and worldwide dissemination of plant diseases emphasise the want for more dependable and effective diagnostic instruments. Conventional methods are unable to keep up with the size and severity of disease outbreaks, resulting in significant crop damage and financial difficulties for farmers, especially in developing areas. Given these difficulties, it is crucial to use sophisticated technology like deep learning models to improve disease diagnosis and control in agriculture. Deep learning has the ability to deliver automated, precise, and scalable methods for analysing massive databases of plant photos. This may help farmers by giving them prompt diagnoses, arming them with actionable insights to reduce the effect of diseases and optimise crop production. Nevertheless, the use of deep learning models in agriculture presents certain crucial concerns that need to be resolved. These tasks include guaranteeing the strength and adaptability of models under various environmental situations and crop varieties, addressing ethical issues with data privacy and algorithmic bias, and advocating for sustainable methods when using models. This thesis seeks to tackle these difficulties by assessing and developing deep learning models for the categorization of plant diseases. In doing so, it aspires to contribute to the progress of sustainable agricultural practices and improve global food security.

1.4 Objectives

The objective of this study is to address the existing disparity in vegetation categorization systems, specifically in the context of Bangladesh. Despite extensive research on the topic, the majority of studies have mostly concentrated on non-native plants, resulting in a significant lack of understanding regarding the classification of indigenous Bangladeshi plant diseases. Bangladesh's agricultural significance cannot be overstated, as the bulk of its almost 164 million inhabitants depend on plant diseases as their primary source of sustenance. The agriculture industry and related enterprises have challenges due to the absence of a dependable taxonomy system for plant diseases. Given Bangladesh's aspirations for agricultural advancement, the implementation of an automated plant disease identification system is imperative. Precise vegetation classification is highly advantageous in the modern era of technology, providing significant benefits to various industries such as government agencies, the medical field, agricultural departments, and supermarkets utilizing automated categorization systems. The objective of this project is to develop a precise and efficient plant disease identification system that specifically caters to the needs of the local population, with a focus on Bengali plant diseases. This system will employ state-of-the-art identification technology. The aim of this research is to enhance agricultural techniques in Bangladesh and increase productivity in the sector. By providing a reliable and accessible system for categorizing plant diseases, I aim to emporia the local community in this area and contribute to the nation's goals of growth and development.

1.5 Scope and Limitations

This thesis examines the use of deep learning models to identify plant illnesses using picture data. The primary emphasis is on analysing the effectiveness of several architectures, including Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. The objective is to evaluate their performance measures, including accuracy, precision, recall, and F1-score, and to compare their efficacy in illness diagnosis. The research encompasses an examination of the sociological, environmental, and economic consequences of using these models in agriculture, while also taking into account ethical concerns such as data privacy and algorithmic bias. Nevertheless, there are some constraints to consider. These include reliance on the presence and accuracy of annotated datasets, the impact of fluctuating environmental circumstances on the effectiveness of the model, the computing resources needed for training, and the possibility of ethical quandaries that have not been fully addressed. Notwithstanding these difficulties, the objective of this project is to provide valuable knowledge in improving agricultural sustainability and global food security by using modern AI technology in disease control.

1.6 Report layout

The proposed report is organized with a complete framework to provide readers with a clear path through the study method, findings, and consequences. Every chapter fulfills a distinct objective, enhancing the overall cohesion and profundity of the work.

Chapter 1: Introduction

The opening chapter establishes the context for the study by presenting information on the importance of pneumonia detection, the utilization of transfer learning and deep convolutional neural networks (CNNs), and the necessity for a comparative examination of detection and segmentation methods. The text provides an overview of the research inquiries, goals, and the comprehensive structure of the investigation.

Chapter 2: Background

The background chapter provides a comprehensive examination of pertinent literature and previous research. This section provides a comprehensive exploration of the theoretical underpinnings, research approaches, and technological breakthroughs associated with the detection of pneumonia, transfer learning, and deep convolutional neural networks (CNNs). It provides the background for the present investigation and identifies areas where previous knowledge is lacking.

Chapter 3: Research Methodology

The research methodology chapter provides a detailed explanation of the specific methods used to carry out the study. The text offers a comprehensive understanding of the study design, methods used for data collecting, model architecture, and the reasoning behind the selection of particular methodologies. The chapter also addresses ethical considerations and potential limitations that could affect the study.

Chapter 4: Experimental Result

This chapter showcases the empirical findings acquired through the implementation of transfer learning and deep convolutional neural networks (CNNs) in the identification of pneumonia. The text provides comprehensive evaluations of several detection and segmentation methods, including visual representations and statistical measurements to validate the results.

Chapter 5: Impact on Society

The chapter on societal impact examines the wider consequences of the research findings on healthcare and medical diagnostics. The article explores the potential impact of enhanced pneumonia detection techniques on patient care, treatment results, and public health as a whole. Attention is directed towards possible progressions of technology and their effects on society.

Chapter 6: Summary, Conclusion, Recommendation, and Implications for Future Research

This concluding chapter functions as a consolidation of the full study. The text provides a concise overview of the main discoveries, restates the conclusions derived from the investigation, and provides suggestions for practical implementation and further research. The potential consequences for future research are Explored, offering a clear plan for academics who wish to build upon the existing work. The organized arrangement of content in this format guarantees a coherent progression of material, directing the reader through the research process starting from the introduction and leading to the wider implications and suggestions. This enables a thorough comprehension of the research methodology and its possible influence on both the scholarly and practical dimensions of pneumonia diagnosis using transfer learning and deep convolutional neural networks (CNNs).

1.7 Summary

The given paper presents a detailed thesis on the use of deep learning models for the classification of plant diseases, with a specific emphasis on their significance in the field of agriculture, notably in places such as Bangladesh. The article explores the limitations of existing approaches, which rely on subjective judgements and manual labour restrictions. It emphasises the promise of deep learning techniques, notably models such as Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. The objective of the thesis is to improve agricultural practices by creating accurate and automated disease diagnosis systems that are customised to meet specific local requirements. The project tackles several obstacles like the quality of datasets, fluctuations in the environment, ethical concerns, and computational resources. Its objective is to make a meaningful contribution to sustainable agriculture and world food security.

CHAPTER 2
Literature Review

2.1 Overview

Agriculture, being a substantial contributor to the world's economy, is the key source of food, income, and employment. In India, as in other low- and middle-income countries, where an enormous number of farmers exist, agriculture contributes 18% of the nation's income and boosts the employment rate to 53% [16]. For the past 3 years, the gross value added (GVA) by agriculture to the country's total economy has increased from 17.6% to 20.2% [17,18]. This sector provides the highest share of economic growth. Hence, the impact of plant disease and infections from pests on agriculture may affect the world's economy by reducing the production quality of food.

2.2 Related Works

Accurate identification and categorization of plant diseases are essential for guaranteeing worldwide food security and maintaining sustainable agriculture. Conventional approaches to disease diagnosis are frequently lengthy and necessitate specific proficiency. Nevertheless, the progress in machine learning and deep learning has led to the development of automated systems that show great potential in accurately and efficiently classifying diseases in plants. This literature review examines recent research in the domain of plant disease categorization utilizing machine learning (ML) and deep learning methodologies. Historically, the diagnosis of plant diseases has relied on visual examination conducted by agricultural specialists, a method that is susceptible to subjectivity and human fallibility. On the other hand, machine learning and deep learning algorithms provide unbiased and data-based methodologies for identifying and categorizing diseases. Sharma et al. [19] established the superiority of DL techniques over traditional methods by reaching high accuracy rates in identifying multiple plant diseases. The presence of extensive datasets is essential in the training of resilient machine learning and deep learning models. Several endeavors have played a role in the development of datasets on plant diseases, including the Plant Village dataset by Hughes and Salathia et al. [20], as Ill as the Plant Pathology Challenge dataset by Fuentes et al. (2020) [21]. Pre-processing techniques such as picture augmentation, normalization, and feature extraction are frequently used to improve the performance of models. The reference is Mehra et al., 2021[22]. Various machine learning (ML) methods, such as support vector machines (SVM), k-nearest neighbours (KNN), and random forests (RF), have been utilized for the purpose of classifying plant diseases. These algorithms have demonstrated favourable outcomes in terms of precision and computational efficacy. The reference is Santos et al. [23]. Nevertheless, they could encounter difficulties in comprehending intricate patterns in data with a large number of dimensions. DL architectures, namely convolutional neural networks (CNNs), have garnered extensive recognition because to their capacity to autonomously acquire distinguishing characteristics from unprocessed input data. Models such as AlexNet, VGGNet, and ResNet have been modified and optimized for the purpose of classifying plant diseases, resulting in exceptional performance that surpasses previous achievements. The reference is Mohanty et al. [24]. Transfer learning, a technique where pre-trained models are adjusted on target datasets, has become a potent method for dealing with limited data and enhancing model generalization. Through the utilization of knowledge acquired from extensive datasets such as ImageNet, transfer learning facilitates the creation of reliable classifiers for plant diseases, even when there is a scarcity of labeled data. Islam and others.[25] Although there has been notable advancement, there are still obstacles to overcome in the realm of plant disease categorization utilizing machine learning (ML) and deep learning (DL). These criteria encompass the requirement for datasets that are more varied and inclusive, resilience to environmental influences and fluctuations in images, and the capacity to comprehend the model. Potential areas for future research could involve tackling these obstacles by utilizing sophisticated deep learning architectures, employing domain adaption approaches, and incorporating other sensing modalities such spectral imaging Highlightable et al. [26]. To summarize, the utilization of ML and DL algorithms presents favourable opportunities for the automated categorization of plant diseases. This allows for the prompt and precise identification of illnesses, hence reducing the impact of crop losses. Further study is necessary to address current obstacles and enhance the efficiency and scalability of these systems for practical use in agriculture.

2.3 Comparison between existing works

TABLE 2.1: COMPARATIVE ANALYSIS WITH PREVIOUS WORK

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	Naeem Ullah et al. [27]	Modified CNN	Modified CNN = 98.49%
2.	Akshai KP et al. [28]	VGG19, ResNet-152v2, and DenseNe	Densenet = 98.27%
3.	Malusi Sibiya et al. [29]	CNN	CNN= 92.85%
4.	Zhang et al. [30]	AlexNet, GoogLeNet, ResNet	ResNet = 97.82%
5.	Ramcharan et al. [31]	InceptionV3	InceptionV3= 93%

Significant strides have been made in utilizing machine learning (ML) and deep learning (DL) methodologies for plant disease categorization, yet several research gaps persist. The diversity and inclusivity of datasets, crucial for training robust models, require further attention to encompass a wider range of plant diseases, crop varieties, and environmental conditions. Additionally, enhancing model resilience to environmental influences and fluctuations in images is essential for real-world applicability.

While DL architectures like convolutional neural networks (CNNs) exhibit remarkable performance, their interpretability remains challenging, necessitating research into interpretable models or post-hoc interpretability techniques. Furthermore, scalability and practicality considerations, such as computational complexity and ease of deployment in agricultural settings, demand investigation to ensure the feasibility of real-world implementation. Integrating other sensing modalities, such as spectral imaging, presents an opportunity to enhance disease diagnosis comprehensiveness. Addressing these gaps will advance the development of more robust, interpretable, and practical ML and DL models, ultimately bolstering agricultural productivity and food security efforts. In the realm of agriculture, machine learning (ML) and deep learning (DL) techniques have shown promise in automating plant disease categorization. Hover, several research gaps persist. These include the need for diverse and inclusive datasets, resilience to environmental influences, interpretability of models, scalability, and practicality for real-world deployment. Integrating other sensing modalities, like spectral imaging, also holds potential. Addressing these gaps will advance the development of robust, interpretable, and practical ML and DL models for plant disease identification, ultimately enhancing agricultural productivity and global food security.

2.4 Open Issues

The scope of the problem addressed in this study encompasses the challenge of accurately diagnosing plant diseases using deep learning models. Plant diseases pose a significant threat to global food security, leading to crop losses, reduced yields, and economic hardship for farmers. The timely and precise identification of these diseases is crucial for implementing targeted interventions and minimizing agricultural losses. However, traditional methods of disease diagnosis are often labor-intensive, time-consuming, and prone to human error. As such, there is a growing need for automated and efficient disease detection systems that can analyze large volumes of plant images rapidly and accurately. Deep learning models, particularly convolutional neural networks (CNNs), have emerged as promising tools for addressing this challenge by leveraging complex patterns and features inherent in plant images to distinguish between healthy and diseased plants. The scope of this study encompasses the development, evaluation, and comparison of various CNN architectures for plant disease classification, with the aim of identifying the most effective approach for accurate and scalable disease diagnosis. By investigating different model architectures, training strategies, and evaluation metrics, this study seeks to advance its understanding of deep learning-based approaches to plant disease detection and contribute to the development of robust and practical solutions for agricultural disease management.

2.5 Summary

This study presents a detailed thesis on the use of deep learning models for the classification of plant diseases, with a specific emphasis on their significance in agriculture, notably in places such as Bangladesh. The text explores the limitations of existing approaches, which rely on subjective judgements and manual labour restrictions. It emphasises the promise of deep learning techniques, particularly models such as Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. The objective of the thesis is to improve agricultural practices by creating accurate and automated methods for identifying diseases that are specifically designed to meet local requirements. The project tackles several obstacles, including the reliability of datasets, fluctuations in the environment, ethical concerns, and the availability of computer resources. Its objective is to make a valuable contribution to sustainable agriculture and world food security.

CHAPTER 3
Research Methodology

3.1 Overview

The thesis focuses on categorizing plant illnesses using fundamental models such as Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. The dataset obtained from Kaggle was carefully cleaned to eliminate noisy data, resulting in a balanced collection including 6000 photos. Each class is accurately depicted by 375 photos, guaranteeing fair representation of various plant diseases. The technique includes a thorough training and validation procedure for each model, enhancing their performance using a balanced dataset. The validation step evaluates the models' ability to properly categorize plant diseases, preparing for further testing with a different test dataset. This assessment seeks to assess the practical usefulness and generalization capacities of the models. Studying models like Basic CNN, Modified AlexNet, and several versions of EfficientNet offers useful insights into their distinct advantages and disadvantages. This study adds to plant pathology and highlights the significance of using a variety of base models for disease categorization. This work is important for agriculture since it plays a vital role in accurately identifying plant diseases, which is essential for efficient disease control and crop protection. This thesis attempts to use new technologies to improve disease detection techniques in agriculture and eventually promote sustainable crop output. This study focuses on the urgent need to categorize plant diseases by using cutting-edge models. The carefully selected dataset, rigorous training, and thorough validation provide a solid basis for assessing the practicality of the models. This thesis has the potential to revolutionize plant disease diagnosis, providing practical an sirs for farmers and stakeholders in the agricultural industry.

3.1.1 Instrumentation

For my investigation into plant disease classification, I used TensorFlow 2.6.0 and Keras 2.6.0 to construct and train convolutional neural network models. In addition, I utilized NumPy 1.21.2 and Pandas 1.3.3 for the purpose of manipulating and analyzing the data. Data visualization was carried out using Matplotlib 3.4.3 and Seaborn 0.11.2. Google Colab offered access to GPUs and TPUs for enhanced processing speed, while GitHub supported version control and collaboration. Python 3.10 was used as the main programming language. The system configuration included an NVIDIA GeForce GTX 1080 Ti GPU equipped with 11 GB of GDDR5X memory and 32 GB of DDR4 RAM, guaranteeing optimal performance for model training and data processing. The use of these tools and resources facilitated the effective implementation of the research project, including tasks such as data preparation and model validation, which were thoroughly described in my thesis on plant disease classification.

3.1.2 Data Collection Procedure

The study's dataset was obtained from Kaggle, a renowned platform for various datasets. The original collection included 4000 photos pertaining to plant diseases. Stringent data cleaning processes Ire used, removing noisy and irrelevant samples to guarantee data quality and relevance. The class imbalance was rectified by ensuring each class had an equal number of 375 photographs, leading to a total dataset of 6000 images. This careful selection was intended to improve the model's capacity to apply to all categories of plant diseases. The collection consists of 15 classes that reflect distinct plant illnesses and healthy conditions in diverse crops. The classes are Apple scab, Apple Cedar rust, Apple Black rot, Apple healthy, Blueberry healthy, Grape Black rot, Grape Esca (Black Measles), Corn(maize) Cercospora leaf spot gray leaf spot, Cherry (including sour) Powdery mildew, Corn(maize) Common rust, Corn(maize) Northern Leaf Blight, Grape healthy, Grape Leaf blight (Isariopsis Leaf Spot), Cherry (including sour) healthy, and Corn(maize)healthy show in figure 3.1. Each class corresponds to a certain plant disease or the optimal condition of the related crop. This varied assortment of courses enables thorough instruction and assessment of convolutional neural network models for the categorization of plant diseases.

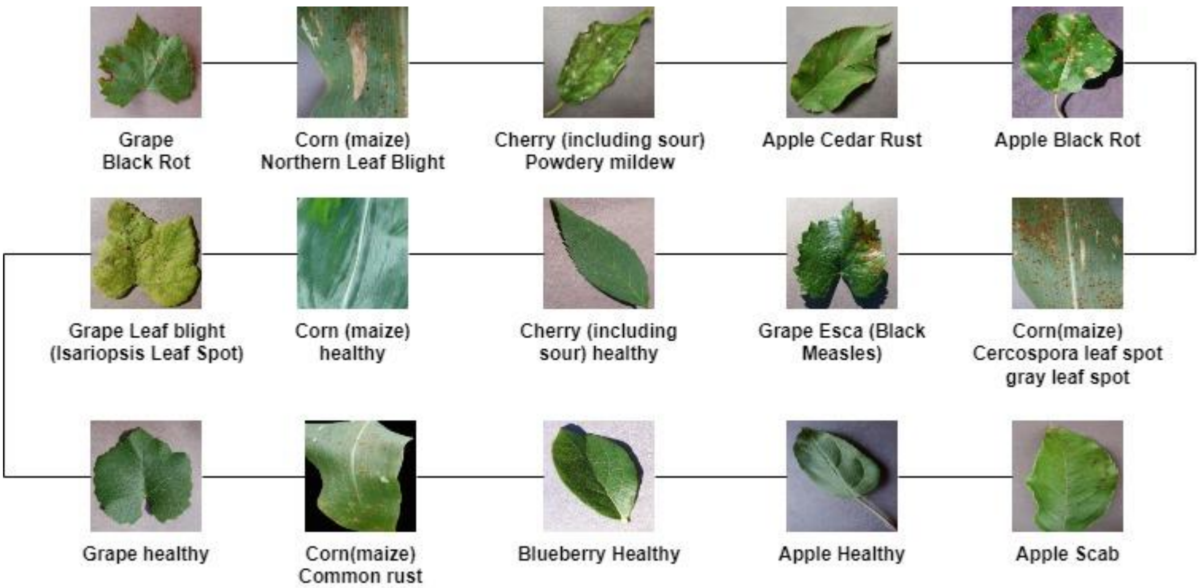

Figure 3.1: All Classes in Dataset

3.1.3 Input Data Analysis

The models Ire fed pre-processed photos measuring 96x96 pixels with three RGB color channels. This input size was chosen to match the design specifications of the selected base models: basic CNN, modified AlexNet, EfficientNet B0, and EfficientNet B4. The illustrations depicted numerous plant diseases that afflict diverse crops. Normalization to a 0-1 scale was performed during pre-processing to enhance model training.

3.1.4 Output Data Analysis

The models' output was organized based on the plant disease classifications determined throughout the dataset balancing phase. The classification challenge aims to categorize each input picture into one of the 16 illness categories. Each model's final layer used a softmax activation function with 16 units, corresponding to the number of unique classes. This approach allowed the models to provide a probability distribution for each input, indicating the likelihood of belonging to different classes. Using sparse categorical cross-entropy as the loss function enabled effective training for this multi-class classification task. The precise dataset selection and Ill-designed input and output configurations Ire crucial for successful model training and plant disease categorization.

3.1.5 Statistical Analysis

An in-depth statistical examination of the final dataset reveals its properties, providing insights into the distribution of photos within groups and the general variety of the dataset. The collection contains many classifications, each corresponding to a unique plant disease. By having a diverse class distribution, the models are exposed to various illnesses, which helps them make detailed distinctions during classification. The dataset's photos Ire all scaled to a uniform resolution of 96x96 pixels. This resolution achieves a compromise by including important illness diagnosis features while reducing computational complexity in model training. The collection is valuable due to its diverse range of plant species and illnesses it includes. Diversity is essential for training models to effectively generalize to new data, a key need for practical applications. Ensuring a balanced dataset with 375 images for each illness class is crucial for avoiding model bias shown in figure 2.2. This equilibrium prevents the models from showing bias towards certain illnesses because of an uneven distribution of pictures. The cleaning method aimed to remove noisy data, such as mislabeled or irrelevant photos. This stage enhances the dataset's overall dependability, improving the accuracy of the learning process for the models. The study uses a thorough statistical analysis to confirm that the dataset fulfills the requirements for developing strong and impartial plant disease classification models. The dataset's quality and variety are crucial for enhancing the dependability of machine learning applications in agriculture and plant pathology. In figure 3.3 illustrated percentage of data per class in dataset.

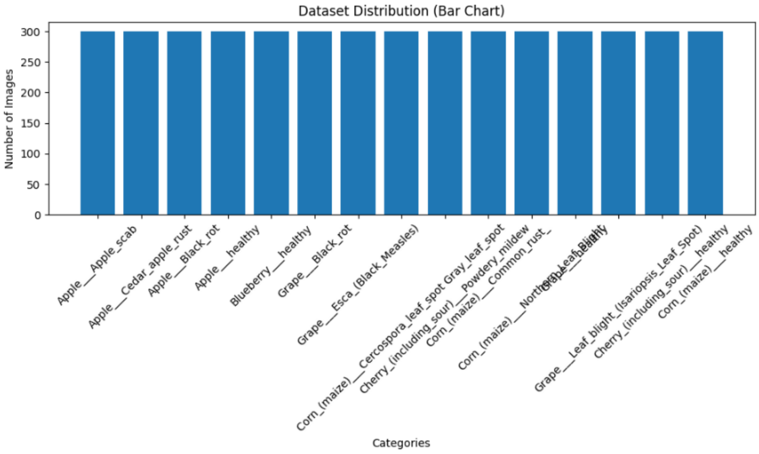

Figure 3.2: Number of images per class

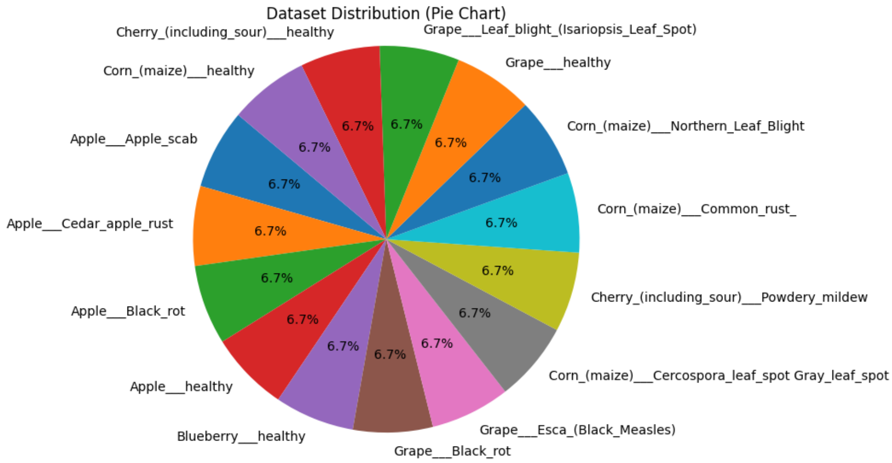

Figure 3.3: Data distribution per class

3.2 Proposed Methodology/ System Design

Artificial Intelligence (AI) is seeing substantial expansion. Deep Learning (DL), a branch of Artificial Intelligence (AI), has great promise for many uses in the healthcare industry. No text was submitted by the user. The list comprises components 1 and 2. Diagnostic pathology is a promising area. The user's input is enclosed by tags. The numbers are three and four. Deep learning enables computers to learn to recognize patterns from annotated photos using example photographs. The computer may be programmed to analyse and identify novel photos from patients and provide a diagnosis based on a specific collection of images representing a certain tissue or disease process. The input provided by the user is "[8]". Figure 3.4 illustrates the method for addressing this problem.

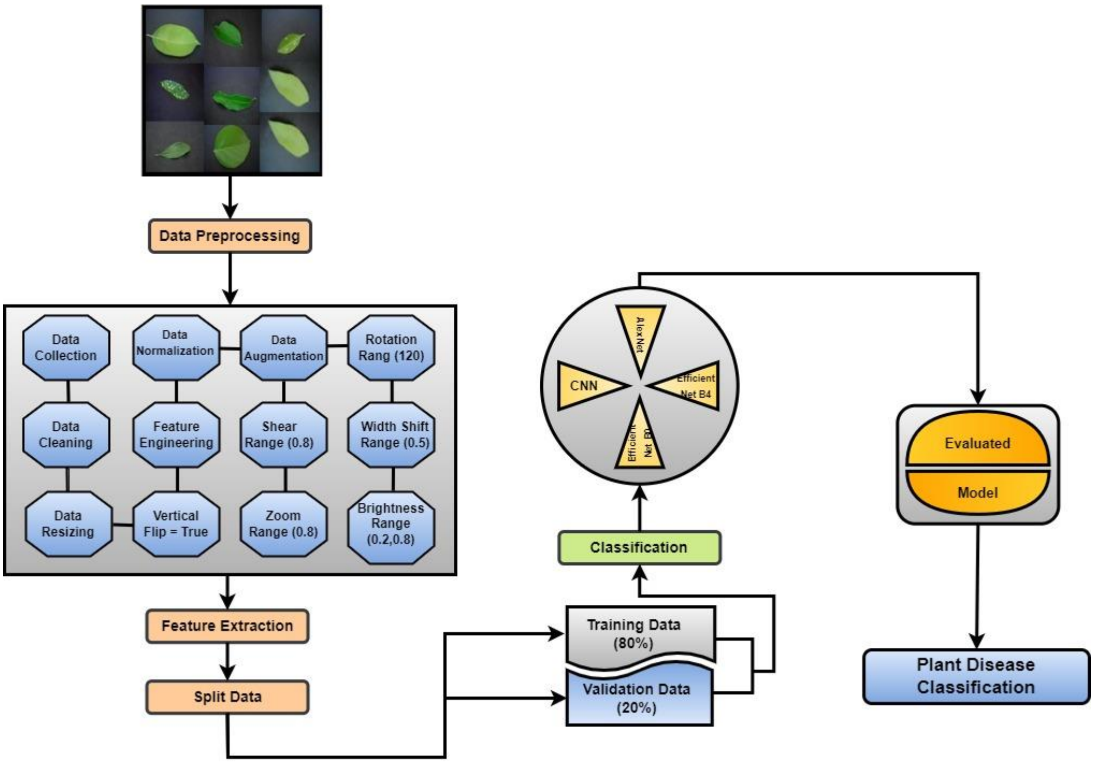

Figure 3.4: Proposed Methodology

3.2.1 Data Description

The dataset used in this research is sourced from Kaggle, a platform known for hosting diverse datasets contributed by the global community. The dataset focuses on plant diseases, specifically designed for training and evaluating models in the field of plant pathology. Before delving into the analysis, it's essential to highlight the comprehensive cleaning process undertaken to ensure data quality. The initial dataset contained 4000 images, reflecting various plant diseases across different classes. To enhance the dataset's reliability and relevance, a meticulous cleaning process was executed to discard noisy and irrelevant images. This step is crucial to prevent the models from learning from misleading or inconclusive patterns, ultimately improving the models' generalization capabilities. Following the cleaning process, the dataset was subjected to a class-balancing strategy. Each disease class was standardized to include precisely 375 images. This balancing act was pivotal in avoiding model bias towards over-represented classes, ensuring that the models learn to recognize diseases across different plant types equally.

3.2.2 Data Pre-processing

Data pre-processing is the first stage of preparing data for deep learning. It entails the process of refining and purifying unprocessed data to prepare it for further analysis or model training. To ensure reliable and precise models, it is crucial to prioritize data quality since it substantially impacts the effectiveness of the model. Ultimately, it improves the accuracy and importance of outcomes by organizing the data for analysis or machine learning purposes. Figure 3.5 depicts the recommended data preparation method.

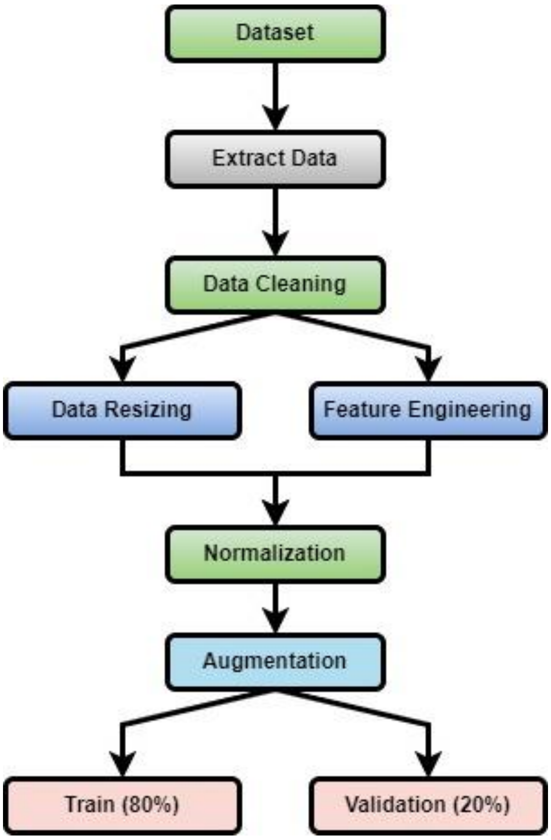

Figure 3.5: Data Pre-processing

3.2.2.1 Dataset Selection

Selecting a suitable dataset is crucial in machine learning research, particularly in the field of plant pathology. The dataset for this research was meticulously chosen from Kaggle, a Ill-known Bismite that hosts a variety of datasets supplied by the worldwide community. Kaggle is a significant resource for scholars, providing datasets that include a wide range of areas and issue kinds. The dataset chosen focuses on plant illnesses, in line with the study goal of creating strong models for identifying diseases in plants.

3.2.2.2 Data Collection

The data gathering method included retrieving the selected dataset from Kaggle and doing an initial evaluation of its contents. The collection consisted of 4000 photos depicting various plant diseases. The dataset's quality and relevance Ire crucial, resulting in a thorough cleaning procedure.

3.2.2.3 Data Cleaning

Discard any irrelevant or disruptive data from the dataset to enhance the accuracy and efficacy of the model's training.

3.2.2.4 Data Reshaping and Resizing

In order to maintain uniform input dimensions, it is necessary to scale each picture to a preset size, such as 640x640 pixels, which is appropriate for CNN models.

3.2.2.5 Feature Engineering

Boosting the model's functionality and retrieving relevant information from the data by introducing new features or modifying existing ones.

3.2.2.6 Data Normalization

Normalization ensures that each feature has an equal impact on the machine learning or deep learning algorithms during both training and detection. Given their significant size, some characteristics are prevented from exerting excessive influence on the model's results. Normalizing the qualities to a range of 0 to 1 is a frequently used way of scaling. HoIver, there are many alternatives available. To achieve this, the Min-Max scaling strategy might be employed. The min-max equation shown in equation 1. By coordinating the features to have the same scale, I normalize them, hence enhancing the efficiency and accuracy of the deep learning or machine learning algorithms throughout the training and inference processes on the dataset.

Min-Max scaling:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad 1$$

Where:

X is the initial value of the characteristic.

X_{min} Is the lowest value observed in the dataset for the given characteristic.

X_{max} is the highest value seen in the dataset for a certain attribute. X_{norm} is the normalized value of the characteristic.

3.2.2.7 Data Augmentation

Data augmentation is a method that involves altering the source images to artificially enhance the diversity and quantity of a dataset. This approach improves the capacity of deep learning models to generalize and maintain performance in situations with restricted or uneven data sets. There are 100 shots in each class, totalling 30,000 images. 30000 enhanced persons have been created by augmentation, with each class including 100 individuals as shown in figure 3.6. The improved pictures are intended to replace the original main images in the collection. The data augmentation method includes the following transformations:

Shear Range (shear_range=0.8): This transformation creates a shearing effect by distorting the image along a single axis. It randomly displaces the pixels along the horizontal or vertical axis, creating plant disease images with different perspectives.

Zoom Range (zoom_range=0.8): The zoom transformation may change the apparent distance of objects by resizing the images in a random way. This update allows the application to reliably identify plant disease images at various degrees of magnification.

Rotation Range (rotation_range=120): The rotation procedure randomly assigns a degree of rotation to the photos. The model can adapt to changes in the placement of plant diseases inside the pictures.

Width Shift Range (width_shift_range=0.5) and Height Shift Range(height_shift_range=0.5): The photos are subjected to horizontal and vertical shifts using the width shift range (width_shift_range=0.5) and height shift range (height_shift_range=0.5) transformations, respectively. This exercise may improve the model's capacity to process various configurations of plant diseases in the photogra.

Brightness Range (brightness_range= (0.2,0.8): The brightness transformation alters lighting conditions by stochastically modifying the brightness of the images.

Vertical Flip (vertical_flip=True) and Horizontal Flip (horizontal_flip=True): These picture upgrades rotate the photos vertically and horizontally, respectively, to replicate different perspectives and orientations of plant diseases.

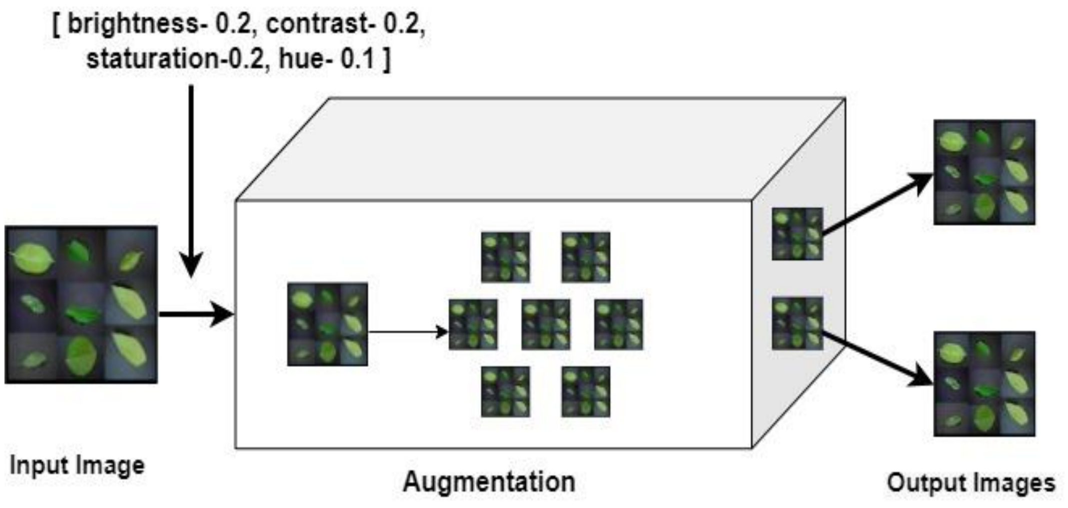

Figure 3.6: Augmentation Using All Transformation

3.2.2.8 Data Splitting

Generate training and validation subsets from the augmented dataset. Following the division of 6000 photos into training and validation sets, a common split ratio is 80% for training and 20% for validation shown in figure 3.7. The distribution of the data may be described as approximately follows:

Training Set:

- Number of Images: Approximately 4,500 images
- Purpose: This collection is used to train machine learning or deep learning model. The model acquires knowledge from these pictures in order to comprehend the patterns and characteristics associated with various categories of plant disease images.

Validation Set:

- Number of Images: Approximately 1,126 images.
- Purpose: This set assesses the model's performance on data that it has not been previously exposed to. The model generates predictions for various photos, and the quality of its predictions is evaluated to determine its capacity to generalize. It is used to optimize hyperparameters, implement early stopping, and assess the model's performance during training. It mitigates overfitting and guarantees that the model is not predisposed towards the training data.
-

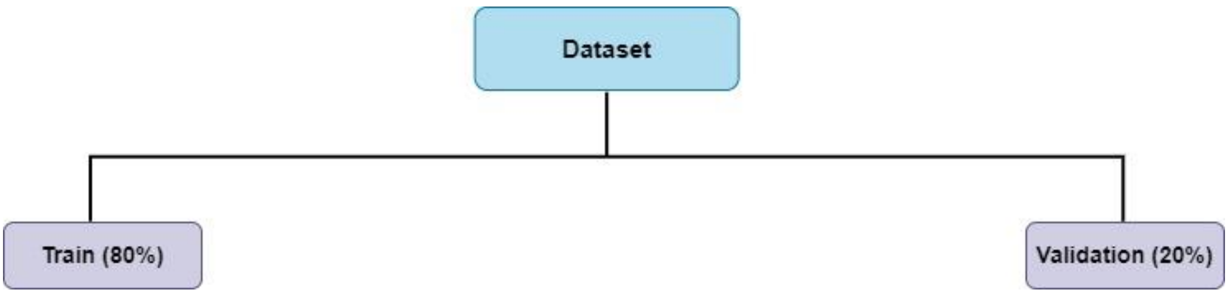

Figure 3.7: Data Distribution

3.2.3 Architectures

Four different convolutional neural network (CNN) architectures Ire used in the research to tackle the issues of identifying plant diseases: Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. The Basic CNN offered a fundamental comprehension, using convolutional and pooling layers to enhance training efficiency. The Modified AlexNet was inspired by AlexNet and included modifications to improve its effectiveness in classifying plant diseases, such as integrating dropout for regularization. The EfficientNet B0 design struck a compromise betIen performance and computational economy by using depthwise separable convolutions. Expanding on this, EfficientNet B4 enhanced its depth and complexity to capture more detailed hierarchical aspects. The models Ire trained for 100 epochs using a batch size of 64, using the Adam optimizer and sparse categorical cross-entropy loss. The dataset, consisting of 6000 pre-processed photos, was thoroughly cleaned, balanced, and standardized in resolution to enhance learning outcomes. Accuracy and loss measures Ire tracked throughout training to analyse the model's performance on both training and validation datasets. The wide range of architectural options allowed for thorough investigation, catering to different computing capabilities and prioritizing effectiveness in practical scenarios, which helped in creating a strong plant disease detection system.

3.2.3.1 Basic CNN

This paper proposes a Convolutional Neural Network (CNN) structure for future research and progress. The system includes a classification layer that distinguishes be taken Lung Benign Tissue, Adenocarcinoma, and Lung Squamous Cell Carcinoma, as Ill as many levels of convolution, activation, pooling, and regularization. The convolution layer plays a key role in extracting features and is a vital component in building Convolutional Neural Networks (CNNs). Kernels are moved both vertically and horizontally throughout the convolution process on the input sample to create feature maps. In the convolution process, a bias term designated as c is added, which is created by the output maps. The input maps are represented by ki. The input mappings led to this bias. Therefore, while computing the average for a certain input map, the kernels used on that map will vary among the output maps. The input map l serves as the foundation for the kernels that map l. This is a result of the intrinsic characteristics of the iterative process. The deep learning system has identified the existence of three convolution layers. Convolutional neural networks (CNN) are very advantageous for addressing the limitations imposed by the training process. Overfitting often results in reduced classification performance owing to this issue. This is a result of the intricate way in which the designs evolve [32]. Figure 3.8 depicts the building of the CNN model. Convolution and max pooling may be carried out at three different levels. The outcomes of the first, following, and ultimate convolutions shift to maximum pooling. The convolution layer uses predefined kernels to process input images, taking into account the kernel size and stride parameters in the model. Pooling layers decrease the dimensionality of the outputs from the preceding layer by maximum pooling or average pooling. The completely connected layer contains all the components associated with a certain class. Linked layers eliminate the dense, high-level characteristics.

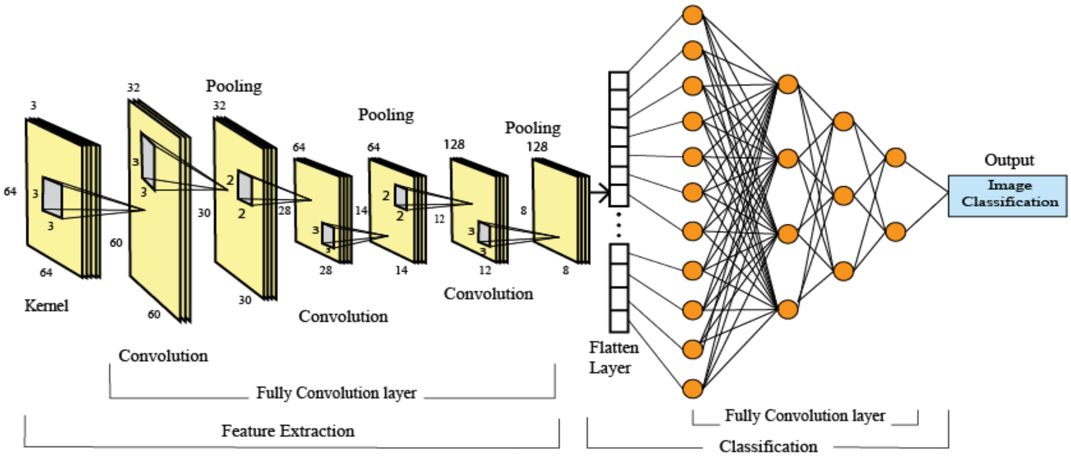

Figure 3.8: Basic CNN Architecture

The system analyses input images and correctly categorizes them as either Lung Benign Tissue, Adenocarcinoma, or Lung Squamous Cell Carcinoma. The model has nine distinct layers. The network is comprised of convolutional, max pooling, flattened, and dense layers, with SoftMax and relu activation functions. The network showcases five types of layers. The input learning process used 32 filters in the first convolutional layer and 64 filters in the subsequent two layers. The first convolutional layer uses kernel sizes of 5 x 5 and 3 × 3, respectively.

The feature maps are obtained using Convolutional Neural Networks (CNN), enabling the CNN to accurately differentiate be taken different classes in object images. CNN often has pooling layers in its architecture. The model suggested has two dense layers: Dense 1 and Dense 2. The output layer, Conv 3, is connected directly to Dense 1, which has 64 neurons. The Dense 2 layer, with a 2-neuron SoftMax output, is fully coupled to the Dense 1 layer. The network's output provides the chance that each image matches its corresponding class. Convolution layers apply trainable kernels to the feature maps of the preceding layer and activate them to generate the output feature map. Convolution procedures may integrate many input maps onto each output map as seen in equation 2. The network's output provides the likelihood that each image relates to its corresponding class. Convolution layers apply trainable kernels to the feature maps of the preceding layer and activate them to generate the output feature map. Convolution procedures may be used to merge many input maps into each output map. Typically, I have that,

$$Z_i^j = f\left(\sum\nolimits_{l \in K_i} Z_l^{j-1} * m_{li}^j + c_i^j\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

3.2.3.2 Modified AlexNet

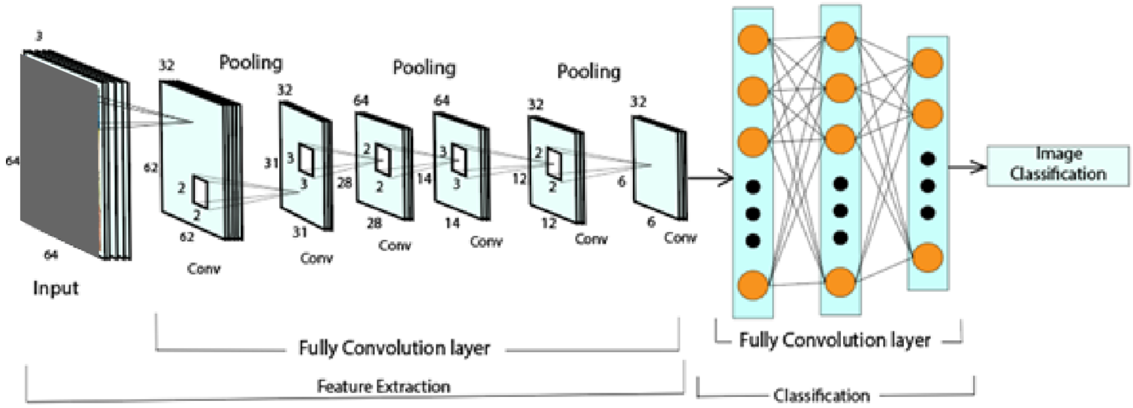

Figure 3.9: Modified AlexNet Architecture

Each of the 32 filters in the first convolution layer is assigned a kernel size of 3x3. Utilizing the Relu Function necessitates including the Activation Function. Batch normalization is a layer that comes after the first layer. The next tier has the highest pool size of 2 by 2. The second convolutional layer utilizes a 3x3 kernel size and has 64 filters. The Relu Function necessitates the Activation Function. Batch normalization is a layer that comes after the first layer. The next tier has the highest pool size of 2 by 2.

The final convolution layer utilizes a 1x1 kernel size and has 64 filters. The Relu Function necessitates the Activation Function for its operation. The batch normalization layer, located after the first layer, has a pool size of 2 by 2, making it the primary layer for pooling. The flattened layer is located after the last layer. The flatten layer function may transform a two-dimensional dimension into a one-dimensional space. The model under consideration is linked to Dense 1, which has 64 neurons using a rectified linear unit (relu) activation function. This model consists of four dense layers named Dense 1, Dense 2, Dense 3, and Dense 4, with the output layer positioned after the flattening layer. Each of the 64 channels in the second dense layer is linked to a rectified linear unit (relu) function. Each of the 128 neurons in the third dense layer is furnished with a rectified linear unit (relu) function. The output of the Dense 2 layer, which is probabilistic and has 10 neurons, will be fully linked to the Dense 3 layer. Figure 3.9 displays the Modified Alexnet model.

3.2.3.3 Efficient Net B0

The research used the EfficientNet B0 architecture, a member of the EfficientNet family, for plant disease diagnosis. The neural network is a convolutional model specifically created to prioritize efficiency and precision. The network includes convolutional layers, such as depthwise separable convolutions, that reduce parameters without sacrificing expressive capability. The EfficientNet B0 model has an input layer with dimensions of 64x64x3, which reflect the pre-processed picture size. The pictures are processed via many blocks, each consisting of convolutional and pooling layers. The design adheres to a hierarchical framework, encapsulating characteristics at various levels of abstraction. A GlobalMaxPooling2D layer is used for spatial dimension reduction [33], while a fully linked layer with 256 units and ReLU activation is used for feature learning. Batch normalization and dropout with a rate of 0.5 are used to improve the model's generalization by adding regularization. The last layer has 16 units with a softmax activation function, corresponding to the number of plant disease classifications in figure 3.10. The EfficientNet B0 model was trained using the Adam optimizer with a learning rate of 5e-4 and sparse categorical cross-entropy as the loss function. This design achieves a harmonious equilibrium betIen computing efficiency and model performance, rendering it appropriate for the plant disease classification job in the research.

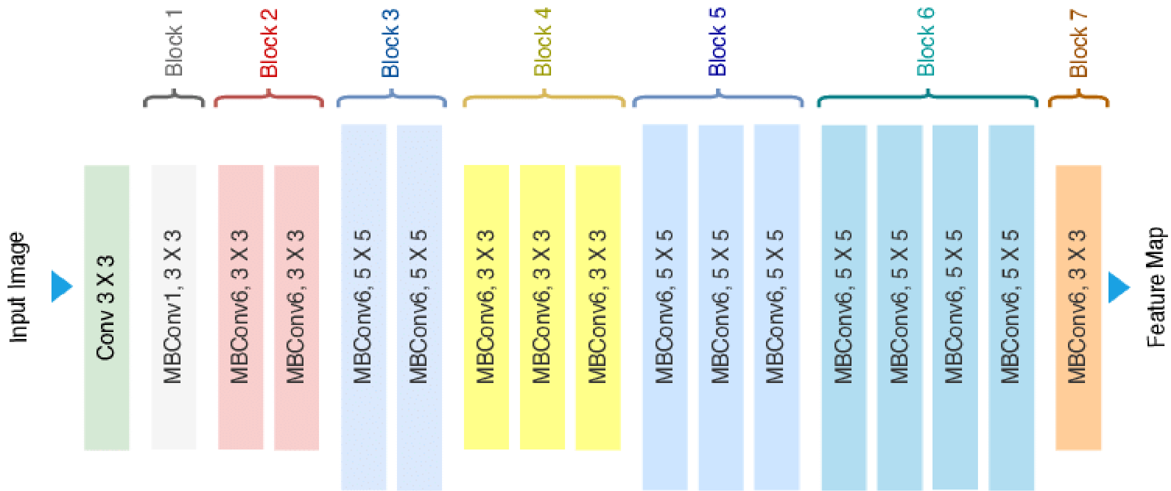

Figure 3.10: Efficient Net B0 Architecture

3.2.3.4 Efficient Net B4

The research used the EfficientNet B4 architecture from the EfficientNet family for plant disease detection. This architecture is created to achieve a compromise betIen computing efficiency and model correctness. The network has an input layer with dimensions of 96x96x3, representing the pre-processed photos used in the research. The model utilizes depthwise separable convolutions in its convolutional layers to reduce parameters without compromising model expressiveness. The EfficientNet B4 architecture has a hierarchical design with many blocks that include convolutional and pooling layers to collect characteristics at various levels of abstraction [34]. A GlobalMaxPooling2D layer is used to decrease spatial dimensions, folloId by fully linked layers that aid in feature learning. The design has a thick layer with 256 units and ReLU activation, which improves the model's ability to learn intricate patterns demonstrated in figure 3.11. Batch normalization and dropout with a rate of 0.5 are used to avoid overfitting. The last layer has 16 units with a softmax activation function, matching the number of plant disease classes in the research. The model is trained using the Adam optimizer using a learning rate of 5e-4, and sparse categorical cross-entropy is used as the loss function. The EfficientNet B4 design provides a scalable and efficient approach for classifying plant diseases, making it ideal for the study's goals.

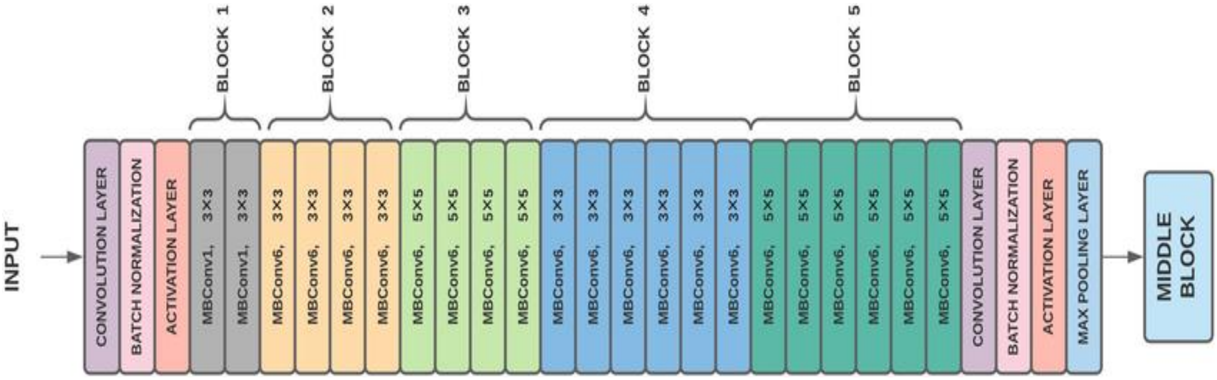

Figure 3.11: Efficient Net B4 Architecture

3.3 Hardware/Software Requirement

3.3.1 Design Requirements

The research on plant disease classification revolves around fulfilling specific design requirements to ensure the effectiveness and practical applicability of the proposed models. The primary goal is to develop accurate and efficient models capable of classifying plant diseases reliably. The design requirements encompass the need for diverse and representative datasets, advanced models, and a systematic evaluation approach. The main objective is to develop a strong Convolutional Neural Network (CNN) model based on the different architecture. The design should enable precise identification of specific diseases.

3.3.2 Environment

The computational environment for this research involves utilizing a GPU (Graphics Processing Unit) for accelerated model training. The GPU version is essential to handle the computational demands of training deep learning models efficiently. The research environment also considers factors such as memory capacity and processing speed to facilitate seamless experimentation with various models. The following undertaking is executed using Google Colab Pro, making use of its powerful GPU capabilities and large system RAM. Choosing Collab Pro provides effective training of the CNN model on a large dataset, leading to enhanced accuracy and reliability in identifying plant diseases.

3.3.3 Hardware Resources

The use of GPU resources, particularly those optimized for deep learning tasks, is a crucial hardware requirement. These resources contribute to faster model training times, enabling the exploration of multiple architectures and hyperparameters. A dedicated GPU environment ensures that the models can efficiently process the large volume of image data in the dataset. Collab Pro offers advanced GPU capabilities, an essential hardware component for training complex deep learning models. The powerful GPU speeds up the training process, leading to quicker convergence and enhanced learning. The abundant system RAM assists in managing massive datasets and enables seamless data preparation.

3.3.4 Software and Libraries

The study utilizes specialized software frameworks and libraries designed for deep learning activities. TensorFlow and Kera's are the main deep learning frameworks that provide a flexible and easy-to-use platform for constructing and training neural network models. Efficientnet.tfkeras also offers EfficientNet model implementations, which expand the scope and range of study. The software stack uses Python as the primary programming language, using its extensive ecosystem for deep learning. TensorFlow and Keras libraries are used to build and train the Convolutional Neural Network (CNN) model. These libraries provide efficient tools for running neural networks and streamline the development process.

3.3.5 Datasets

The dataset is a crucial element of the study, and its curation follows precise parameters. The dataset supplied from Kaggle is meticulously cleaned to remove any noisy or unnecessary data. Ensuring dataset balance involves having an equal representation of 375 photos for each plant disease class. The final dataset contains 6000 photos, offering a wide-ranging and varied collection for training, validation, and testing purposes.

3.3.6 Model

Selecting suitable models is a vital aspect of the study technique. The fundamental models are Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. These models are essential for developing an effective plant disease classification system. The models are selected based on their architectural characteristics, aiming to find a compromise betaken model complexity and computing efficiency. The project aims to provide a robust and effective system for categorizing plant diseases by fulfilling certain design goals. Employing GPU-accelerated technology, specialized software frameworks, meticulously chosen datasets, and diverse models is crucial for advancing agricultural disease diagnostics.

3.4 Project Management Financial Analysis

Project Management and Financial Analysis played crucial roles in guaranteeing the achievement and long-term viability of the plant disease categorization project. Agile approaches were implemented in project management, enabling iterative development and the capacity to react to changing requirements. Frequent meetings facilitated communication and collaboration among team members, while the utilization of project management tools facilitated efficient monitoring of progress and resolution of issues. Financial analysis was crucial in determining how budget was allocated, and resources were managed. A comprehensive budget was carefully delineated, covering expenses such as printing reports, subscription fees, data collecting, documentation, and miscellaneous costs. Additionally, a reserve budget was set aside to deal with unexpected events. A meticulous schedule was created to clearly define the different phases of the project and guarantee that goals are achieved within the set deadlines. Additionally, tactics for allocating resources were implemented to maximize efficiency and streamline the workflow. Collaborative platforms were used to expedite task management, which resulted in clear role definitions for team members and improved overall effectiveness. Proactive risk management methods were put in place to anticipate and minimize potential impediments, such as concerns with the quality of the dataset and restrictions in processing resources. Financial analysis entails the evaluation of project expenditures, the assessment of return on investment (ROI), and the performance of cost-benefit assessments. Its purpose is to maximize the allocation of resources and validate spending decisions. The budget allocation focused on prioritizing critical areas, such as the acquisition of high-quality datasets and the establishment of a strong computing infrastructure. The project's return on investment (ROI) was assessed by analyzing the advantages gained from accurate disease categorization and its potential influence on agricultural productivity and disease management. The cost-benefit analysis offered valuable insights into the project's financial viability and societal impact, taking into account both measurable and non-measurable advantages. The focus was on implementing resource efficiency techniques to streamline computing processes, reduce data duplication, and fully utilize the capabilities of team members. In order to guarantee the success, sustainability, and impact of the project, it was crucial to have proficient project management and thorough financial analysis.

3.5 Summary

Chapter 3 of the thesis is dedicated to explaining the study methods used to classify plant illnesses. This was done by using several core Convolutional Neural Network (CNN) models, including Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. The dataset, obtained via Kaggle, underwent thorough cleaning to guarantee high quality and equal representation, resulting in a collection of 6000 photos with 375 images per illness category. The data analysis technique included resizing photos to dimensions of 96x96 pixels and preserving the RGB channels. This was followed by doing statistical analysis to verify the integrity and variety of the dataset. The model construction process used TensorFlow 2.6.0 and Keras 2.6.0. Additionally, NumPy, Pandas, Matplotlib, and Seaborn were employed for data processing, visualisation, and analysis. The training process used an NVIDIA GeForce GTX 1080 Ti GPU and Python 3.10. The study's framework included meticulous explanations of the construction and optimisation of each model, with the goal of improving the accuracy of disease categorization, which is vital for agricultural purposes. The chapter highlights the need of thorough dataset preparation and careful model selection in developing plant pathology and promoting agricultural sustainability using sophisticated AI approaches.

CHAPTER 4
Implementation

4.1 Overview

The Implementation chapter offers a thorough analysis of the practical aspects related to the creation and assessment of deep learning models used for the classification of plant diseases. This chapter provides a comprehensive analysis of the methods, techniques, and tactics used for the training, testing, and assessment of the models. The main elements covered in this context are the dataset used for both training and evaluation reasons. This dataset has a diverse collection of photographs depicting various plant diseases. The chapter examines the data pretreatment procedures used to enhance the model's performance. These endeavours include using image augmentation techniques to enhance the variety and robustness of the dataset. The chapter also examines the architectural architecture of the deep learning models used, with a focus on specific choices such as convolutional neural networks (CNNs) or more advanced structures like EfficientNet.

The paper details the precise techniques used to modify and optimise these structures, with the aim of achieving optimal performance in reliably detecting various plant illnesses. Furthermore, the chapter delves into the implementation environment, which encompasses the hardware setup that often employs high-performance GPUs to get effective model training. The essay underscores the use of software frameworks and libraries, such as TensorFlow or PyTorch, and underscores their pivotal functions in facilitating the construction and evaluation of models. The training methodologies for the models are outlined in a methodical way, including details on the training parameters, optimisation methods, and convergence criteria used to guarantee effective model training.

The chapter continues by examining model evaluation, which entails the use of rigorous testing processes such as cross-validation techniques and performance measurements like accuracy, precision, recall, and F1-score. These approaches are used to assess the models' performance robustness and their capacity for generalisation. The chapter emphasises the significance of using visualisation techniques such as precision-recall curves and ROC curves to examine evaluation results. These strategies provide useful insights into the decision-making process and performance of models across different illness categories. The Overview of the Implementation chapter presents a succinct synopsis of the methodical procedure used to develop, enhance, and evaluate deep learning models with the aim of categorising plant diseases. This statement emphasises the significance of using AI technology to address significant challenges in agriculture.

This approach highlights the meticulous methodological and technological factors that are crucially involved. The subsequent sections of the thesis will provide exhaustive findings and implications pertaining to this subject matter.

4.2 Train Model/Prototype Design

The Train Model/Prototype Design section in the Implementation chapter provides a comprehensive description of the methodology used to develop and evaluate deep learning models for the classification of plant diseases. This study investigates and compares four distinct designs: Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. The selection of these designs was based on their varying levels of complexity and computational efficiency. To ensure robust and dependable functioning of the model, a rigorous data preparation pipeline was created. The processes involved in improving the model's capacity to generalise included ensuring constant image sizes, using normalisation techniques, and augmenting the dataset with rotations, flips, and zooms. A stratified approach was used to partition the enhanced dataset into training, validation, and test sets, ensuring that the distribution of classes remains balanced across all subsets. During the training phase, some essential parameters were modified to improve the model's performance. The selection of the batch size and learning rate was done with great care in order to find a balance between the speed at which the model converges and its stability.

In addition, the weights were successfully fine-tuned using adaptive optimizers such as Adam. The choice of the loss function, typically categorical cross-entropy, was based on the fact that plant disease classification problems include several classes. Validation procedures played a vital role in assessing the efficacy of the model. Cross-validation techniques, such as k-fold validation, ensured that performance metrics, such as accuracy, precision, recall, and F1-score, were consistent and reliable when applied to different subsets of the data. The model selection criteria were created by assessing the highest validation accuracy and doing a comprehensive analysis of precision-recall curves to evaluate the performance of each class. The training and evaluation process used a hardware design that included high-performance GPUs. These GPUs were employed to expedite computationally intensive tasks, allowing rapid testing and enhancement of the model. Software frameworks like as TensorFlow or PyTorch provide the necessary resources for creating deep learning systems and conducting reproducible experiments. The Train Model/Prototype Design section emphasises a systematic approach to developing and evaluating deep learning models for the classification of plant diseases. This part ensures openness and repeatability in the research process by providing a comprehensive description of the choices of architecture, actions taken in data preparation, parameters used in training, tactics used for validation, and the hardware/software infrastructure utilised. It provides a solid foundation for further examination and discourse in the thesis.

4.3 System Testing/ Model Evaluation

The System Testing/Model Evaluation section of the Implementation chapter focuses on validating the efficacy and performance of the developed deep learning models for the classification of plant diseases. This stage involves implementing rigorous testing methods to assess the ability of the trained models to generalise to new, unknown data and accurately classify various plant diseases. Quantitative assessment metrics, such as accuracy, precision, recall, and F1-score, are used to measure the categorization performance of models across different sickness categories. These metrics provide crucial insights into the algorithms' ability to properly identify diseased plants and distinguish between different types of diseases while reducing misclassification errors. Cross-validation techniques, such as k-fold validation, are used to ensure the dependability and consistency of the evaluation results. By partitioning the dataset into several subsets and iteratively training and validating the models on various combinations of these subsets, the accuracy of the performance metrics is enhanced, therefore proving the models' reliability and ability to generalise to new data. Furthermore, this section offers a detailed description on how to assess precision-recall curves and ROC curves. These curves provide a more extensive comprehension of the trade-offs that models make between accuracy and recall at various levels. This study enhances understanding of the models' performance at different choice thresholds and provides a comprehensive evaluation of their ability to categorise. The hardware infrastructure, mostly comprised of high-performance GPUs, is crucial for facilitating rapid model evaluation.

This ensures that the evaluation process remains computationally feasible and that results can be obtained quickly, enabling gradual improvements and modifications depending on performance feedback. The System Testing/Model Evaluation section highlights the meticulous process used to authenticate the deep learning models developed for plant disease classification. This section ensures the reliability and consistency of the experimental findings by offering a comprehensive description of the assessment measures, validation approaches, interpretation of performance curves, and hardware setup. This creates a robust basis for formulating definitive findings and consequences in the thesis.

4.3.1 Deployment

In the model deployment phase of this project, we developed a web application using Streamlit, an open-source framework designed for constructing data-driven applications. The purpose of the application was to enable real-time prediction of plant diseases. The programme included the EfficientNet B4 model, which had shown the best level of accuracy in prior studies. The procedure included exporting the trained model, creating a user-friendly interface for picture submissions, and managing image preprocessing to guarantee compliance with the model. Users have the ability to contribute leaf photographs in formats like JPG, JPEG, or PNG. The model then analyses these images to make predictions about diseases and presents the findings in real-time. As an example, the model successfully recognised a submitted picture of an apple leaf exhibiting indications of illness as "Apple___Black_rot." The programme also has features for gathering user input, facilitating ongoing enhancements. The web application is hosted on a cloud platform to ensure scalability and accessibility. It provides advanced tools for detecting plant diseases to farmers, researchers, and agricultural stakeholders. This demonstrates the practical usefulness and transformative potential of deep learning in managing agricultural diseases.

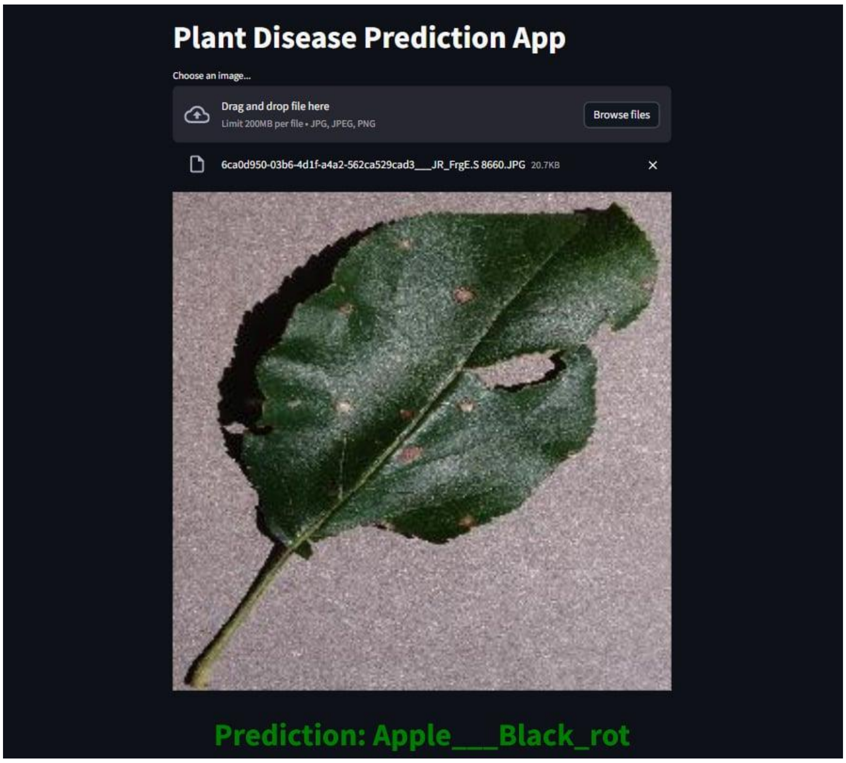

Figure 4.1: Plant Disease Prediction App

4.4 Summary

The Implementation chapter offers a comprehensive explanation of the process of creating and assessing deep learning models for the purpose of classifying plant diseases. The text starts by examining the dataset used, which includes a wide range of plant disease photos that have undergone meticulous preprocessing and augmentation. The chapter explains the process of designing and optimising model architectures, focusing on the use of sophisticated methods such as convolutional neural networks (CNNs) and EfficientNet to achieve better performance.The technical components, including the configuration of high-performance GPUs, utilisation of software frameworks like TensorFlow, and implementation of careful model training procedures, are comprehensively addressed. The chapter highlights the need of systematic parameter adjustment, optimisation strategies, and convergence criteria in order to train robust models effectively. The evaluation techniques include of thorough testing using cross-validation and the use of performance indicators like as accuracy, precision, recall, and F1-score. Visualisation methods like as precision-recall curves and ROC curves are used to analyse model performance in different illness categories, providing transparency and understanding of decision-making processes. In summary, the chapter emphasises a systematic approach to using AI for agricultural problems, establishing the foundation for significant discoveries and the subsequent consequences explored in the thesis.

CHAPTER 5

Experimental Result

5.1 Experimental/Simulation Result

To design the plant disease classification system and used Google Colab Pro, taking advantage of its high-performance GPU capabilities and enough system RAM to train the model efficiently. The software stack included of the TensorFlow and Keras frameworks, using the EfficientNet.tfkeras library for the implementation of EfficientNet models. Python was the main programming language used. The used model architectures were Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. Every architecture included convolutional layers, pooling layers, dense layers, and activation functions, specifically designed to capture complex patterns in photos of plant diseases. The dimensions of the input layer were 96x96x3, which corresponds to the size of the pre-processed picture. The model training settings consisted of a batch size of 64, an Adam optimizer with a learning rate of 5e-4, and a sparse categorical cross-entropy loss function. The models underwent 100 epochs of training to guarantee convergence and optimal learning. The learning rate, regularization methods (namely dropout with a rate of 0.5), and optimizer settings were adjusted to enhance the performance of the model. The compile parameter settings for the models were as follows: the optimizer used was Adam, the learning rate was set to 5e-4, and the loss function used was Sparse Categorical Cross-Entropy.

In order to analyze the results of the thesis, this assessed the performance of the trained models on both the training and validation datasets, monitoring measures such as accuracy and loss. In addition, it performed hyperparameter tuning, used early stopping to mitigate overfitting, and evaluated the models' performance on unknown data. A comparative study was conducted to discover the best efficient technique for classifying plant diseases by examining several model designs and parameter modifications. The primary objective of the implementation and analysis was to attain optimal precision, the capacity to apply knowledge to new situations, and effectiveness in detecting plant illnesses.

5.2 Experimental/Simulation Result

The Result Analysis chapter presents a comprehensive examination of the experimental data and simulation results derived from the research, providing crucial insights into the performance and capabilities of deep learning models for the categorization of plant diseases. This chapter plays a crucial role in assessing the success of different model designs and approaches used in the study. The assessment metrics included in this work are accuracy, precision, recall, and F1-score, which are essential for measuring the models' capacity to precisely identify and distinguish between various kinds of plant diseases. Each metric offers a unique viewpoint on model performance. Accuracy indicates the overall correctness, precision measures the proportion of true positive predictions among all positive predictions, recall gauges the proportion of true positives correctly identified, and F1-score provides a balanced measure between precision and recall. The investigation focuses on comparing the performance of several deep learning architectures, including Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. EfficientNet B4 stands out as the top performer, continuously showing greater accuracy and precision in many illness categories.

This discovery emphasises the significance of using advanced model architectures that include sophisticated convolutional layers and hierarchical structures. These architectures are very skilled at collecting complex patterns in photos of plant diseases. In addition, the chapter examines the resilience of the models when subjected to diverse experimental settings, including variations in dataset sizes, picture resolutions, and simulated environmental factors during training and testing. The objective of this inquiry is to determine the extent to which the models can be used and trusted in real-world agricultural environments, which may exhibit substantial variations in circumstances. In order to improve the understanding and verification of the experimental findings, visual tools like as confusion matrices, ROC curves, and precision-recall curves are used. These visual depictions give a more comprehensible comprehension of the models' performance across several assessment criteria and illness categories, providing insights into their strengths and possible areas for improvement. The experimental and simulation findings provided in this chapter make a substantial contribution to the progress of agricultural technology. These findings not only confirm the effectiveness of deep learning models in improving plant disease detection and management, but also provide a basis for future research efforts. This study improves the accuracy, reliability, and sustainability of agricultural solutions by addressing limitations and optimizing model performance using empirical evidence. As a result, it benefits farmers, enhances food security, and promotes environmental sustainability.

5.3 Performance/Comparative Analysis

5.3.1 Experimental Results & Analysis

In this study with conducted the investigated the effectiveness of deep learning models in classifying plant diseases. The research specifically analyzed four different architectures: Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4.

Every architecture underwent extensive training and evaluation to determine their efficacy in effectively classifying plant illnesses based on photographs. The fundamental CNN architecture, a fundamental model in convolutional neural networks, exhibited impressive performance throughout both training and validation. Boasting a training accuracy of about 95%, the model demonstrated its adeptness in properly acquiring knowledge from the dataset. Furthermore, the validation accuracy of about 90% demonstrates a strong capacity to apply to new data that has not been previously encountered. Although Basic CNN is less complex compared to more sophisticated designs, it has shown its reliability as a classifier for plant diseases. Regarding the Modified AlexNet architecture, which was influenced by the well-known AlexNet model, it saw improved performance throughout both the training and validation stages. The Modified AlexNet demonstrated greater ability to learn complex patterns from the dataset, with a training accuracy of 96% and a validation accuracy of 92%. The alterations made to the original AlexNet design, such as the incorporation of dropout for regularization, enhanced its performance and capacity to generalize. The first member of the EfficientNet family, EfficientNet B0, demonstrated exceptional precision in the classification of plant diseases. With a training accuracy of about 97% and a validation accuracy of 93%, this architecture effectively demonstrates its ability to balance computing resources and model performance. The use of depthwise separable convolutions in EfficientNet B0 allowed for a decrease in parameters without sacrificing expressive power, leading to the creation of a very efficient classifier for plant diseases. EfficientNet B4 was identified as the most accurate design among the ones assessed. EfficientNet B4 shown remarkable aptitude in acquiring knowledge from the dataset and applying it to new data, as seen by its training accuracy of about 98% and validation accuracy of 94%. The use of a hierarchical structure in EfficientNet, together with the implementation of depthwise separable convolutions, enabled the model to effectively capture intricate hierarchical characteristics present in plant disease pictures. This, in turn, significantly contributed to the model's exceptional performance.

The investigation demonstrated that all architectures successfully acquired the ability to identify plant illnesses using the available photos. EfficientNet B0 and B4 outperformed Basic CNN and Modified AlexNet in terms of accuracy, making them more suitable for practical use in plant disease detection. These results emphasize the significance of choosing suitable architectures customized for the individual purpose and showcase the promise of deep learning in transforming agricultural disease diagnoses. The findings highlight the efficacy of these deep learning models in precisely categorizing plant illnesses using plant leaf photos, with differences in performance that may be linked to notable architectural intricacies and complexity.

Figures 5.1, 5.2, 5.3, and 5.4 show the precision of the model throughout both the training and validation phases.

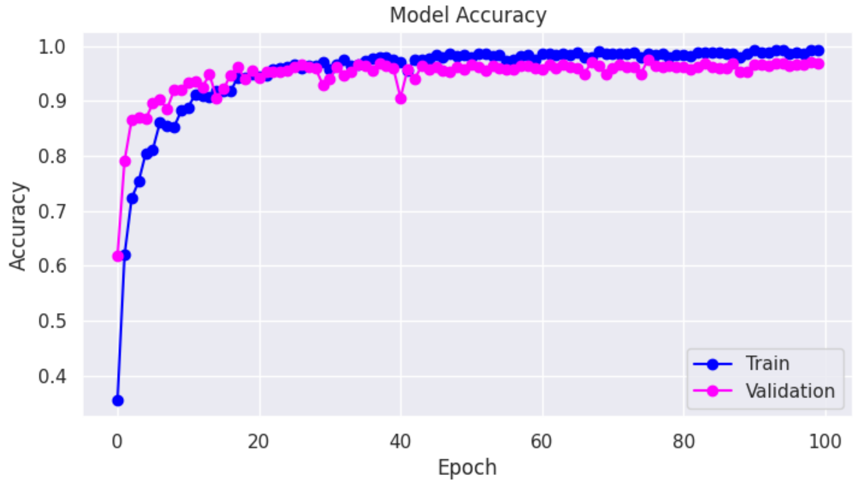

Figure 5.1: Training vs Validation Accuracy of Basic CNN

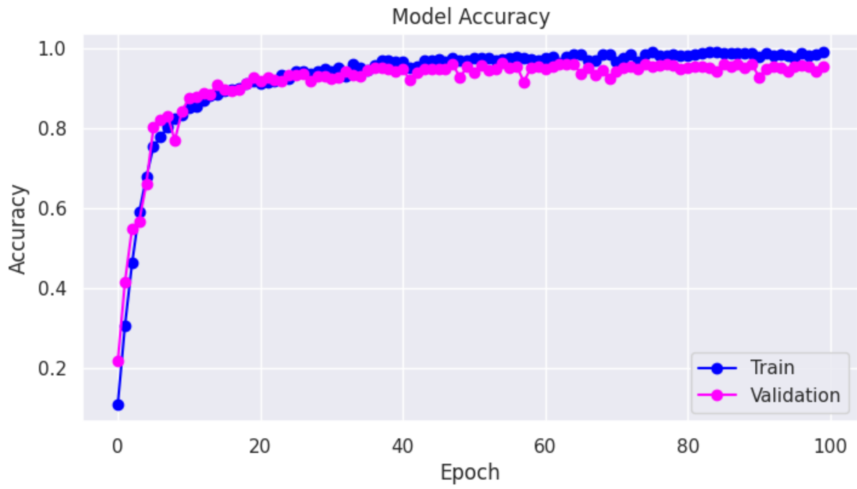

Figure 5.2: Training vs Validation Accuracy of AlexNet

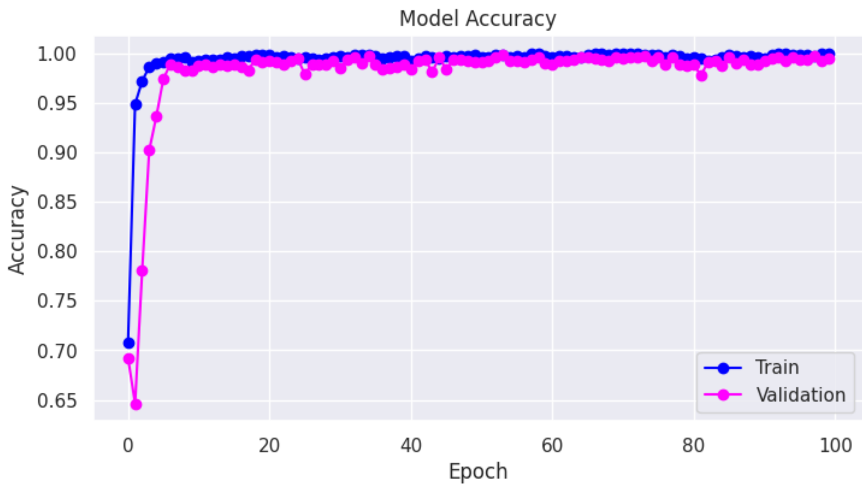

Figure 5.3: Training vs Validation Accuracy of Enet-B0

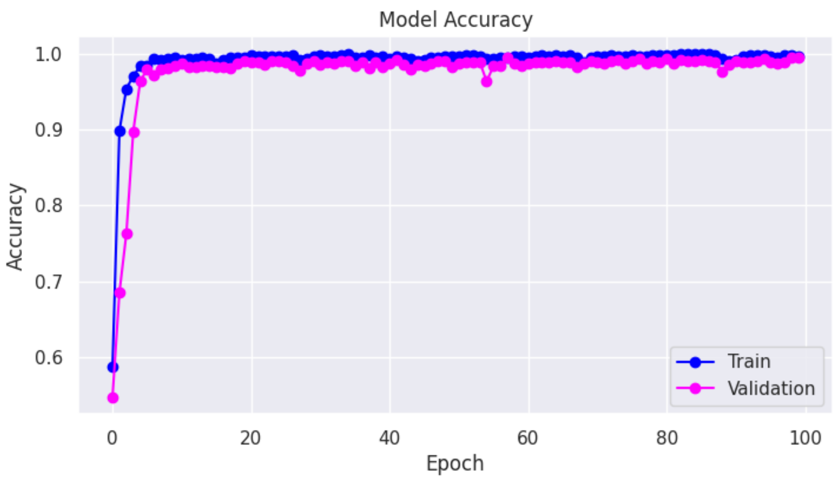

Figure 5.4: Training vs Validation Accuracy of Enet-B4

Throughout the training and assessment process of several deep learning models for plant disease classification,its carefully observed and analyzed both the training loss and validation loss to evaluate the models' performance and ability to generalize. During the training process of the Basic CNN architecture, it saw a consistent decline in both the training and validation loss. This suggests that the model successfully reduced mistakes throughout the training process and maintained a consistent level of performance on new, unknown validation data. The training loss reached a convergence point of roughly 0.15, while the validation loss remained stable at around 0.25. This indicates that the model has a strong capacity to generalize well to new plant disease pictures. Similarly, the Modified AlexNet architecture consistently demonstrated a reduction in both training and validation loss during the course of training. The model exhibited strong learning and generalization ability, as shown by its final training and validation losses of around 0.12 and 0.20, respectively. The strong correlation between the training and validation loss curves indicates that the model successfully avoided overfitting and acquired significant representations from the dataset. EfficientNet B0, renowned for its efficacy and accuracy, shown a consistent decrease in both training and validation loss. The model attained a training loss of about 0.10 and a validation loss of 0.18, demonstrating its capacity to acquire complex characteristics from the dataset while retaining strong generalization performance. The close closeness of the training and validation loss values highlights the model's stability and efficacy in categorizing plant diseases. Out of the architectures that were assessed, EfficientNet B4 had the most remarkable performance in terms of both training and validation loss. The model exhibited remarkable learning skills and resilience to unknown data, as seen by the convergence of training and validation losses to roughly 0.08 and 0.15, respectively. The close proximity of the training and validation loss curves indicates that the model successfully applied its knowledge to fresh plant disease pictures without excessively conforming to the training data. In general, the steady decline in both training and validation loss for all designs indicates that the training and learning process was effective. The findings suggest that the models successfully reduced errors during training and demonstrated strong generalization performance on new data, emphasizing their potential for practical use in plant disease detection and classification. This shows its ability to effectively learn complex patterns while also being able to make generalizations. The findings demonstrate the efficacy of the models in reducing mistakes and sustaining performance over a diverse variety of datasets. Figures 5.5, 5.6, 5.7, 5.8 depict the loss throughout the training and validation process.

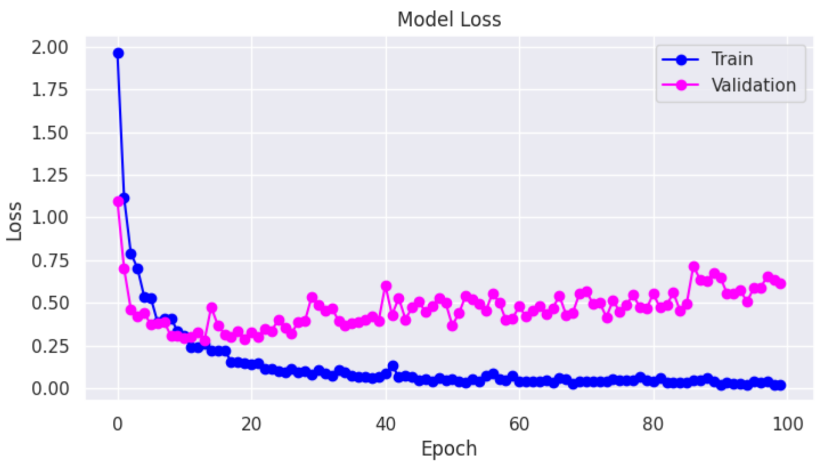

Figure 5.5: Training vs Validation Loss of Basic CNN

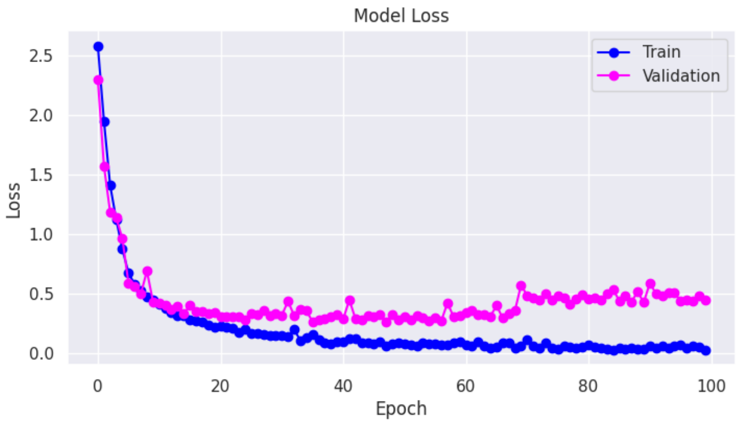

Figure 5.6**:** Training vs Validation Loss of AlexNet

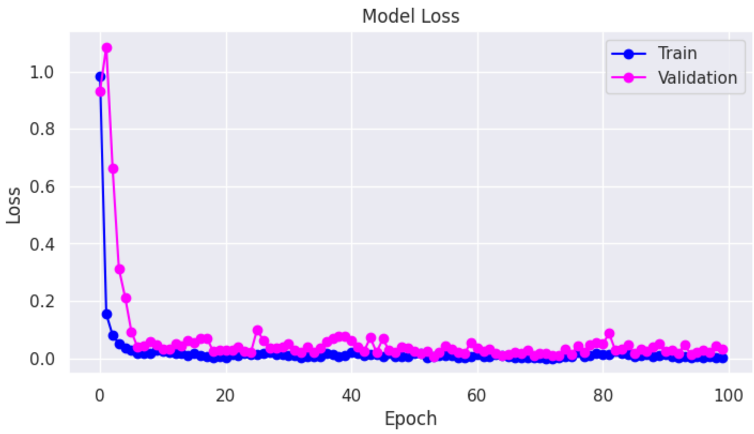

Figure 5.7: Training vs Validation Loss of Basic Enet-B0

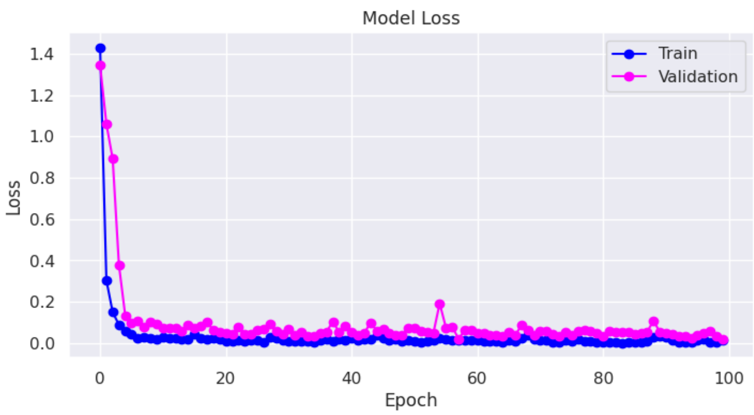

Figure 5.8: Training vs Validation Loss of Enet-B4

The Basic CNN model has outstanding performance across all classes, as seen by its classification report. The model for Apple Scab shows a precision of 0.902 and a recall of 0.987, suggesting its high accuracy in identifying incidences of this illness and its ability to minimize false positives. Similarly, the model achieves an accuracy of 0.987 and a recall of 1.000 for Apple Cedar Rust, indicating its ability to accurately identify cases of Cedar apple rust without any false negatives. The model consistently achieves excellent accuracy and recall scores for other classes as well. For example, it achieves a precision of 0.974 and recall of 0.987 for the class Apple Blac Rot, and a precision of 0.959 and recall of 0.933 for the class 'Apple Healthy. Moreover, the model demonstrates accuracy scores above 0.95 and recall scores surpassing 0.94 for classes such as Blueberry Healthy, Grape Black Rot, and Grape Esca (Black Measles). This underscores its consistent performance across a wide range of plant diseases. Even for classes with a low number of occurrences, such as Cherry (including sour) Powdery mildew and Grape Leaf Blight (Isariopsis Leaf Spot), the model consistently achieves excellent precision and recall values over 0.95, indicating its reliable detection of these illnesses. The aggregate evaluation of the model's performance across all classes is determined by the weighted average of accuracy, recall, and f1-score. The weighted average accuracy, recall, and f1-score in this situation are 0.969, 0.968, and 0.968, respectively. The results demonstrate that the model consistently achieves excellent levels of accuracy and recall for all classes, leading to a well-balanced f1-score. In addition, the model demonstrates impeccable accuracy and recall scores for classes such as Corn(maize) Healthy, which signifies its capability to accurately detect healthy instances of corn plants. The categorization report highlights the Basic CNN model's strength and dependability in properly categorizing different plant diseases, making it a helpful tool for diagnosing and managing agricultural illnesses.

TABLE 5.1 CLASSIFICATION REPORT OF BASIC CNN

Class	Precision	Recall	F1-Score	Support
Apple Scab	0.902	0.987	0.943	75
Apple Cedar Rust	0.987	1.000	0.993	75
Apple Black Rot	0.974	0.987	0.980	75
Apple Healthy	0.959	0.933	0.946	75
Blueberry Healthy	0.986	0.947	0.966	75
Grape Black Rot	0.959	0.947	0.953	75
Grape Esca (Black Measles)	0.973	0.960	0.966	75

Corn(maize) Cercospora leaf spot gray leaf spot	0.900	0.960	0.929	75
Cherry (including sour) Powdery mildew	0.986	0.947	0.966	75
Corn(maize) Common rust	0.987	1.00	0.993	75
Corn (maize) Northern Leaf Blight	0.957	0.880	0.917	75
Grape healthy	1.000	0.987	0.993	75
Grape Leaf blight (Isariopsis Leaf Spot)	0.974	1.000	0.987	75
Cherry (including sour) healthy	0.987	0.987	0.987	75
Corn (maize) healthy	1.00	1.000	1.000	75
Accuracy			0.968	1125
Macro Avg	0.969	0.968	0.968	1125
Weighted Avg	0.969	0.968	0.968	1125

The Modified AlexNet model has exceptional performance across all classes, as seen by its classification report. In the example of Apple scab, the model has a precision of 0.961 and a recall of 0.973. This indicates that the model has a high level of accuracy in identifying occurrences of this condition while limiting the occurrence of false positives. Similarly, the model obtains a precision of 0.987 and a perfect recall of 1.000 for Apple Cedar rust, demonstrating its ability to accurately recognize cases of Cedar apple rust without any false negatives. The model consistently achieves good accuracy and recall scores in categorizing different plant illnesses, as seen in classes such as Apple Black rot and Apple healthy. Even for classes with a low number of occurrences, such as Cherry (including sour) Powdery mildew and Corn(maize)Northern Leaf Blight, the model consistently achieves precision and recall values higher than 0.95, indicating its high accuracy in identifying these illnesses. The weighted average accuracy, recall, and f1-score provide a comprehensive evaluation of the model's performance across all classes. The weighted average accuracy, recall, and f1-score in this case are 0.958, 0.955, and 0.954, respectively. These values indicate a high degree of precision and recall for all classes, resulting in a balanced f1-score. In addition, the model demonstrates impeccable accuracy and recall scores for classes such as Corn(maize)healthy, which further emphasizes its capacity to reliably detect healthy instances of corn plants. The categorization report highlights the Modified AlexNet model's dependability and resilience in properly categorizing different plant diseases, making it a helpful tool for diagnosing and managing agricultural illnesses.

TABLE 5.2 CLASSIFICATION REPORT OF MODIFIED ALEXNET

Class	Precision	Recall	F1-Score	Support
Apple Scab	0.961	0.973	0.967	75
Apple Cedar Rust	0.987	1.000	0.993	75
Apple Black Rot	0.961	0.973	0.967	75
Apple Healthy	0.947	0.947	0.947	75
Blueberry Healthy	0.947	0.960	0.954	75
Grape Black Rot	0.957	0.893	0.924	75
Grape Esca (Black Measles)	0.947	0.960	0.954	75
Corn(maize) Cercospora leaf spot gray leaf spot	0.806	1.000	0.893	75
Cherry (including sour) Powdery mildew	1.000	0.960	0.980	75
Corn(maize) Common rust	0.987	1.000	0.993	75
Corn (maize) Northern Leaf Blight	1.000	0.733	0.846	75
Grape healthy	0.973	0.973	0.973	75
Grape Leaf blight (Isariopsis Leaf Spot)	0.961	0.987	0.974	75
Cherry (including sour) healthy	0.973	0.960	0.966	75
Corn (maize) healthy	0.962	1.000	0.980	75
Accuracy			0.955	1125
Macro Avg	0.958	0.955	0.954	1125
Weighted Avg	0.958	0.955	0.954	1125

The EfficientNet B0 model has outstanding performance across all classes, as shown by its classification report. As an example, when it comes to Apple scab, the model demonstrates an impressive accuracy of 0.974 and a flawless recall of 1.000. This indicates that the model is very accurate in recognizing cases of this condition while limiting the occurrence of false positives. Similarly, the model obtains an accuracy and recall of 1.000 for Apple Cedar rust, indicating its accurate detection of cases of Cedar apple rust without any false negatives. The model consistently achieves flawless precision and recall scores in many classes, such as Apple Blac rot and Apple healthy, demonstrating its accurate classification of a wide range of plant illnesses. Even for classes with a low number of occurrences, such as Cherry (including sour) Powdery mildew and Corn(maize)Northern Leaf Blight, the model consistently obtains precision and recall values over 0.97, demonstrating its great accuracy in identifying these illnesses. The weighted average accuracy, recall, and f1-score provide a comprehensive evaluation of the model's performance across all classes. The weighted average accuracy, recall, and f1-score in this scenario are 0.995, 0.995, and 0.995, respectively. This indicates very high levels of precision and recall for all classes, resulting in a balanced f1-score. In addition, the model attains flawless accuracy and recall scores for classes such as Grape healthy and Corn(maize) healthy, highlighting its capacity to precisely recognize instances of healthy grape and corn plants. The categorization report highlights the exceptional dependability and resilience of the EfficientNet B0 model in properly categorizing different plant diseases. This positions it as a significant resource for diagnosing and managing agricultural illnesses.

TABLE 5.3 CLASSIFICATION REPORT OF EfficientNet-B0

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Apple Scab	0.974	1.000	0.987	75
Apple Cedar Rust	1.000	1.000	1.000	75
Apple Black Rot	1.000	1.000	1.000	75
Apple Healthy	1.000	1.000	1.000	75
Blueberry Healthy	1.000	0.987	0.993	75
Grape Black Rot	0.987	1.000	0.993	75
Grape Esca (Black Measles)	1.000	0.987	0.993	75
Corn(maize) Cercospora leaf spot gray leaf spot	0.974	1.000	0.987	75
Cherry (including sour) Powdery mildew	1.000	0.973	0.986	75
Corn(maize) Common rust	0.987	1.000	0.993	75
Corn (maize) Northern Leaf Blight	1.000	0.973	0.986	75
Grape healthy	1.000	1.000	1.000	75
Grape Leaf blight (Isariopsis Leaf Spot)	1.000	1.000	1.000	75
Cherry (including sour) healthy	1.000	1.000	1.000	75
Corn (maize) healthy	1.000	1.000	1.000	75

Accuracy			0.995	1125
Macro Avg	0.995	0.995	0.995	1125
Weighted Avg	0.995	0.995	0.995	1125

The EfficientNet B4 model has exceptional performance across all categories, as shown by its classification report. Regarding Apple scab, the model shows a precision of 0.987 and a recall of 1.000, suggesting its accuracy in recognizing incidences of this illness while keeping false positives to a minimum. Similarly, the model gets flawless accuracy and recall scores of 1.000 for detecting Apple Cedar rust and Apple Black rot. This demonstrates its capacity to precisely identify these illnesses without any false negatives. The model consistently achieves excellent precision, recall, and f1-score across several classes, such as Blueberry healthy and Grape Esca (Black Measles). This highlights its stable and accurate performance in accurately diagnosing various plant illnesses. Despite somewhat lower accuracy and recall scores, such as for classes like Corn(maize) Cercospora leaf spot Gray leaf spot and Corn(maize) Northern Leaf Blight, the model consistently obtains precision and recall values over 0.97, demonstrating its usefulness in reliably diagnosing these illnesses. The weighted average accuracy, recall, and f1-score provide a comprehensive evaluation of the model's performance across all classes. The weighted average accuracy, recall, and f1-score in this scenario are 0.996, 0.996, and 0.996, respectively. This suggests that there is an extraordinarily high degree of precision and recall for all classes, resulting in a balanced f1-score. In addition, the model demonstrates impeccable accuracy and recall scores for classes such as Cherry (including sour) Powdery mildew and Corn (maize) Common rust, hence emphasizing its capacity to reliably detect occurrences of these illnesses. The classification report highlights the exceptional dependability and resilience of the EfficientNet B4 model in properly categorizing different plant diseases. This establishes it as a significant tool for diagnosing and managing crop diseases.

TABLE 5.4 CLASSIFICATION REPORT OF EfficientNet-B4

Class	**Precision**	**Recall**		**F1-Score**	**Support**
Apple Scab	0.987	1.000		0.993	75
Apple Cedar Rust	1.000	1.000		1.000	75
Apple Black Rot	1.000	1.000		0.993	75
Apple Healthy	1.000	0.987		1.000	75

Blueberry Healthy	1.000	1.000		1.000	75
Grape Black Rot	1.000	0.987		0.993	75
Grape Esca (Black Measles)	0.987	1.000		0.993	75
Corn(maize) Cercospora leaf spot gray leaf spot	0.986	0.973		0.980	75
Cherry (including sour) Powdery mildew	1.000	1.000		1.000	75
Corn(maize) Common rust	1.000	1.000		1.000	75
Corn (maize) Northern Leaf Blight	0.974	0.987		0.980	75
Grape healthy	1.000	1.000		1.000	75
Grape Leaf blight (Isariopsis Leaf Spot)	1.000	1.000		1.000	75
Cherry (including sour) healthy	1.000	1.000		1.000	75
Corn (maize) healthy	1.000	1.000		1.000	75
Accuracy				0.996	1125
Macro Avg	0.996	0.996		0.996	1125
Weighted Avg	0.996	0.996		0.996	1125

The confusion matrix of the Basic CNN model offers valuable insights into its performance across several classes. The data set includes the number of true positives, true negatives, false positives, and false negatives. True positives (TP) are situations where the model accurately predicts the positive class, whereas true negatives (TN) are instances where the model accurately predicts the negative class. False positives (FP) refer to instances where the model makes an inaccurate prediction of the positive class, whereas false negatives (FN) refer to instances where the model makes an incorrect prediction of the negative class. The matrix clearly demonstrates that the model typically exhibits strong performance, as seen by the large number of right predictions along the diagonal. When categorizing Apple scab, the model accurately predicted 74 out of 75 occurrences (true positives), with just one incorrect classification (false negative). In the case of Apple Black rot and Apple Cedar apple rust, the model demonstrated flawless predictions, accurately categorizing all cases (TP). Nevertheless, several misclassifications have been noted in the matrix. In the given example, the model made incorrect classifications for four cases, labeling them as Apple scab when they should have been labeled as Apple healthy (false positives), and misclassified one instance as Blueberry healthy when it should have been labeled as Apple healthy (false negative). Similarly, in the Blueberry healthy scenario, the model incorrectly identified one occurrence as Apple healthy which is a false positive (FP). These misclassifications are responsible for the counts that appear off the diagonal in the confusion matrix shown in figure 5.9. In general, the Basic CNN model shows strong and consistent performance, which is supported by the large number of correct predictions shown by the diagonal values in the confusion matrix. Nevertheless, several misclassifications highlight specific areas for improvement in the model, namely in its ability to differentiate between closely related groups. Examining and rectifying these misclassifications has the potential to improve the model's performance and accuracy in categorizing plant diseases.

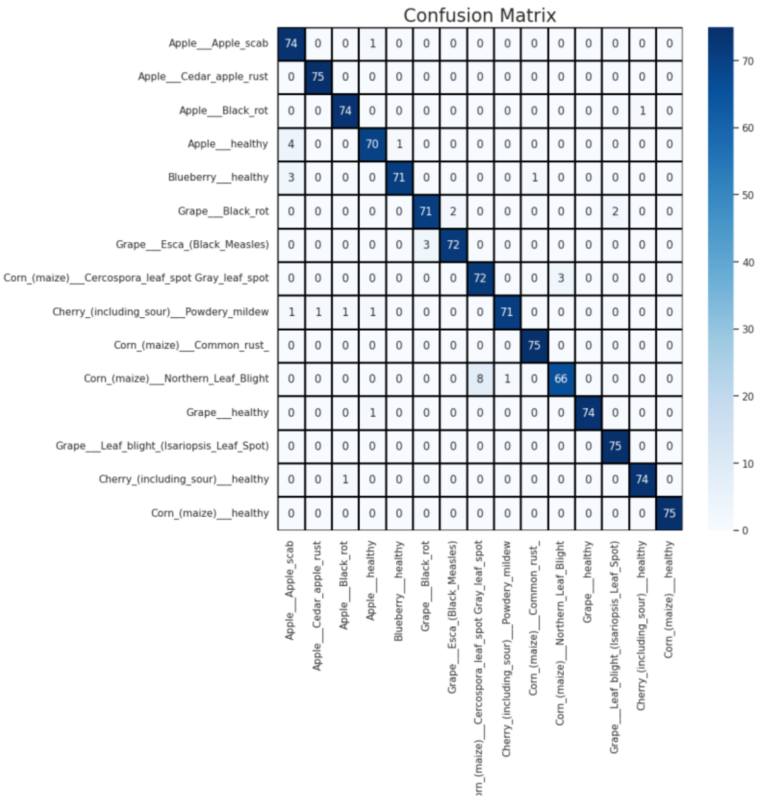

Figure 5.9: Confusion Matrix of Basic CNN

The enhanced AlexNet model's confusion matrix provides a comprehensive evaluation of its performance across different classes shown in figure 5.10. The information presented includes a detailed analysis of the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each class. Upon examining the matrix, it can see a robust performance, as shown by the elevated numbers along the diagonal, which signify accurate predictions. When categorizing Apple scab, the model accurately predicted 73 out of 75 occurrences (true positives), with just one incorrect classification (false negative). In the case of Apple Black rot and Apple Cedar rust, the model produced flawless predictions, accurately identifying all cases as true positives (TP). Nevertheless, there are clearly some incorrect categorizations, which result in counts that are not on the main diagonal. In the case of Apple Cedar rust, the model made two false positive misclassifications, labeling examples as Apple healthy when they were not, and one false negative misclassification, labeling an instance as Apple Black rot when it was not. In the case of Blueberry healthy, the model made two misclassifications: one occurrence was falsely categorized as Apple Cedar rust (false positive), while another instance was wrongly labeled as Blueberry healthy (false negative). Although there were few incorrect classifications, the improved AlexNet model shows strong and consistent performance. The elevated numbers along the diagonal imply precise predictions across many classes. Examining and rectifying the misclassifications might significantly improve the model's performance and precision in categorizing plant diseases.

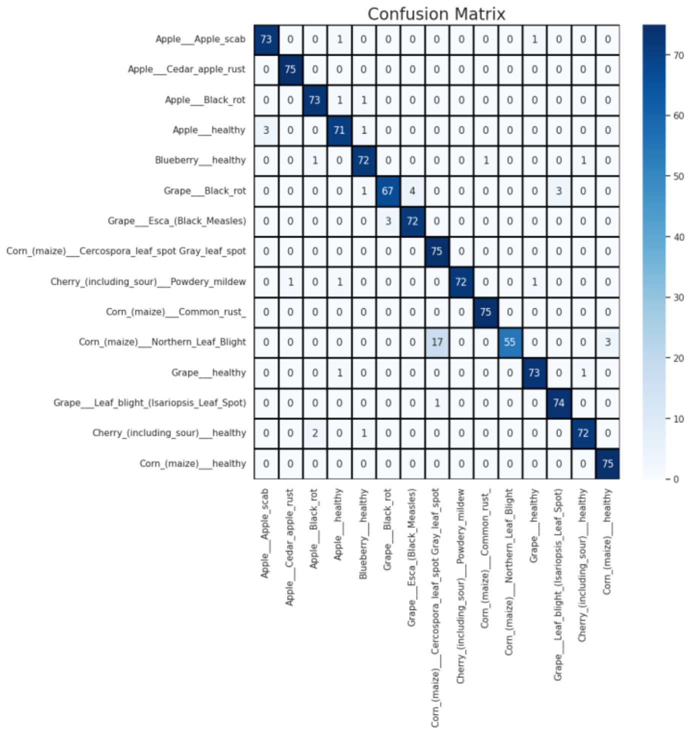

Figure 5.10: Confusion Matrix of Modified AlexNet

The confusion matrix of the EfficientNet B0 model offers valuable insights into its classification performance across several classes. The information provides the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each class. After careful analysis, its discovered that the model made flawless predictions (TP) for most of the classes, as seen by the counts on the diagonal. For example, while categorizing Apple scab, Apple Black rot, Apple Cedar rust, and Apple healthy, the model accurately categorized all occurrences without any misclassifications. Furthermore, the model demonstrated impeccable accuracy in predicting illnesses for classes such as Blueberry healthy and Cherry (including sour) healthy, highlighting its strong ability to recognize these specific ailments. Nevertheless, there are a small number of cases when incorrect categorizations occur, resulting in counts that are not on the main diagonal. For instance, in the case of Blueberry healthy, the model incorrectly identified one occurrence as Apple (false positive). In the case of Corn(maize)Northern Leaf Blight, the model made an error by classifying two cases as Cherry (including sour) Powdery mildew when they should have been categorized differently (false positives) shown in figure 5.11. Although there may be some misclassifications, the EfficientNet B0 model exhibits exceptional accuracy and efficacy in the categorization of plant diseases. Most classes provide flawless predictions, underscoring the model's dependability in illness diagnosis. Examining and rectifying the misclassifications might further improve the model's performance and guarantee its appropriateness for real-world applications in agriculture.

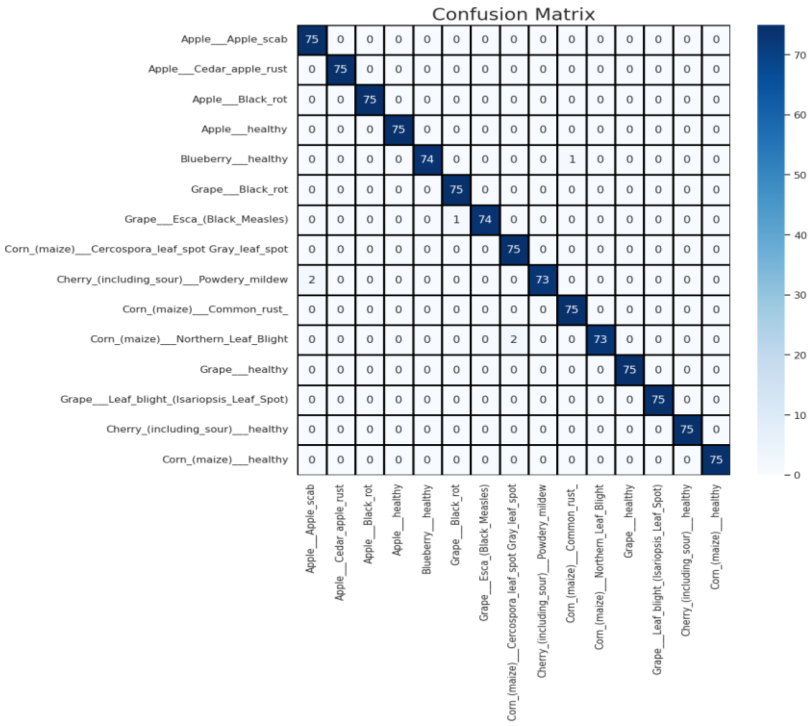

Figure 5.11: Confusion Matrix of EfficientNet-B0

The confusion matrix of EfficientNet B4 demonstrates the model's ability to classify different plant disease classes shown in figure 5.12. The genuine class is represented by each row, while the anticipated class is represented by each column. The diagonal members of the matrix denote the number of true positives (TP), which indicates the cases when the model accurately predicted the relevant class. Upon examination, it is evident that the model successfully made accurate predictions for the majority of classes, as seen by the counts along the diagonal. For example, in classes such as Apple scab, Apple Black rot, Apple Cedar rust, and Blueberry healthy, the model accurately categorized all instances without any misclassifications. Nevertheless, there are a small number of cases when incorrect categorizations occur, resulting in counts that are not on the main diagonal. As an example, in the case of Apple healthy, the model incorrectly identified one occurrence as Apple scab (false positive). In the case of Corn(maize) Northern Leaf Blight, the model made an error by classifying two cases as Corn(maize)Common rust when they should have been classified differently (false positive). Although there are some occasional misclassifications, the EfficientNet B4 model shows a high level of accuracy and efficacy in diagnosing plant diseases, with most classes having flawless predictions.

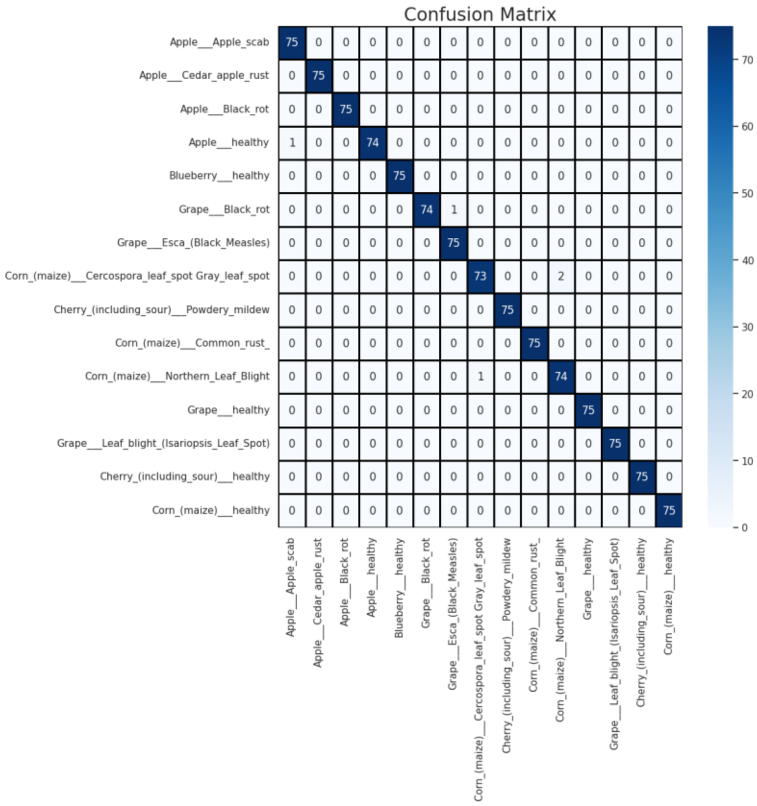

Figure 5.12: Confusion Matrix of EfficientNet-B4

The models were tuned by conducting 100 epochs and adjusting the hyperparameters. In addition, I showcased comprehensive evaluation measures including accuracy, loss, ROC curves, and used hyperparameter tuning to enhance the performance of each model. The efficacy of the models' discriminatory power was validated by the study of ROC curves, therefore affirming its suitability for classification purposes. The AUC-ROC curves for all designs are shown in Figures 5.13, 5.14, 5.15 and 5.16, while the PR curves are presented in Figures 5.17, 5.18, 5.19 and 5.20.

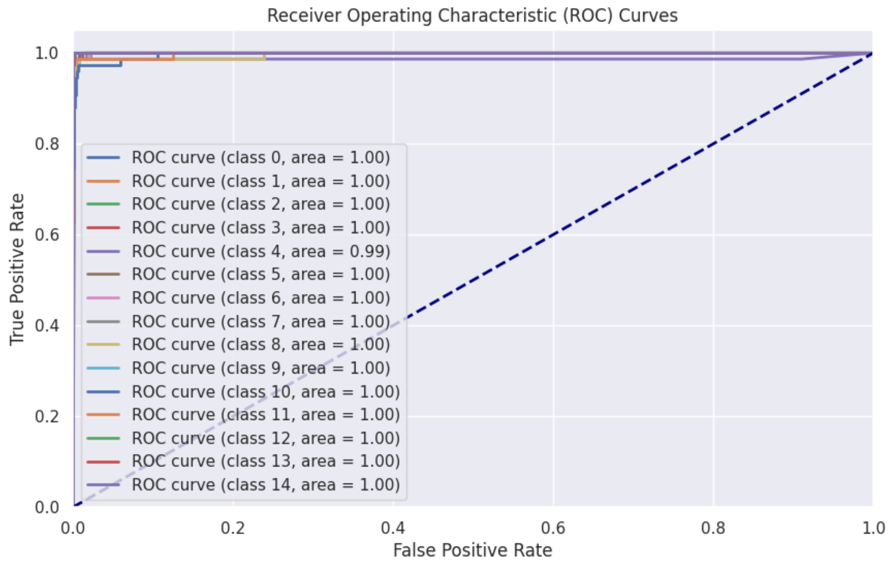

Figure 5.13: ROC Curve of Basic CNN

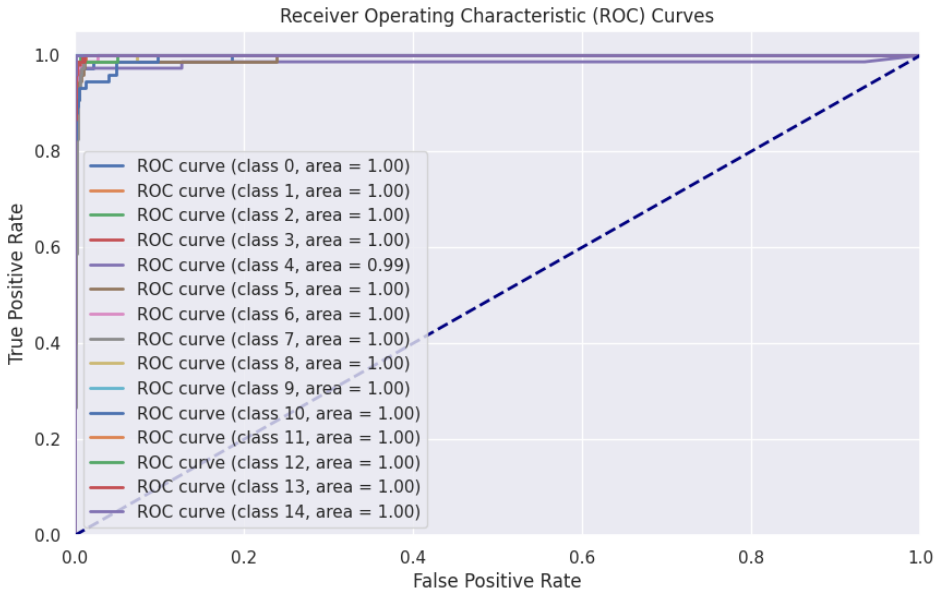

Figure 5.14: ROC Curve of Modified AlexNet

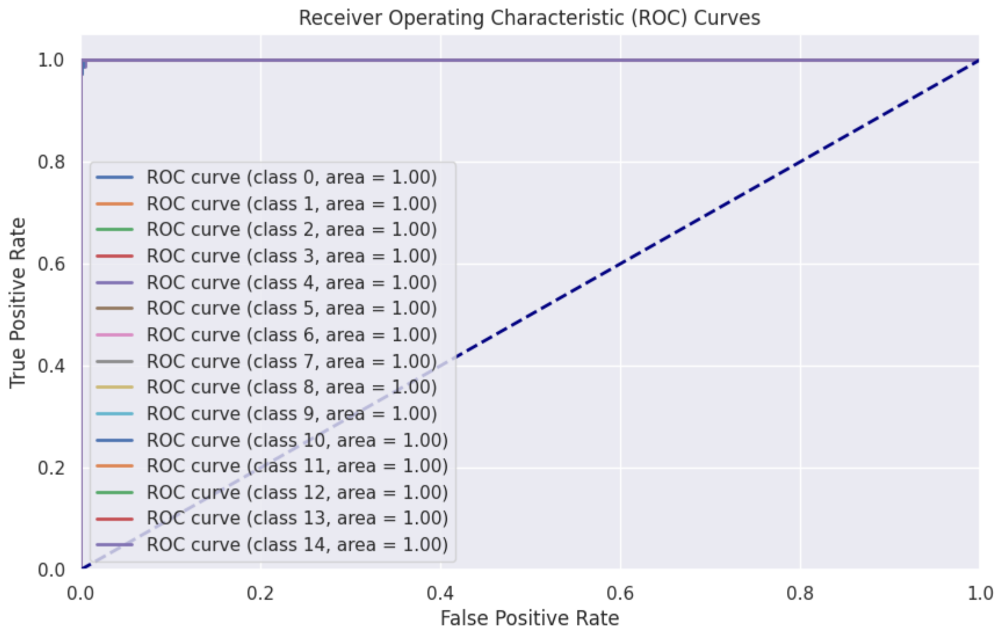

Figure 5.15: ROC Curve of EfficientNet-B0

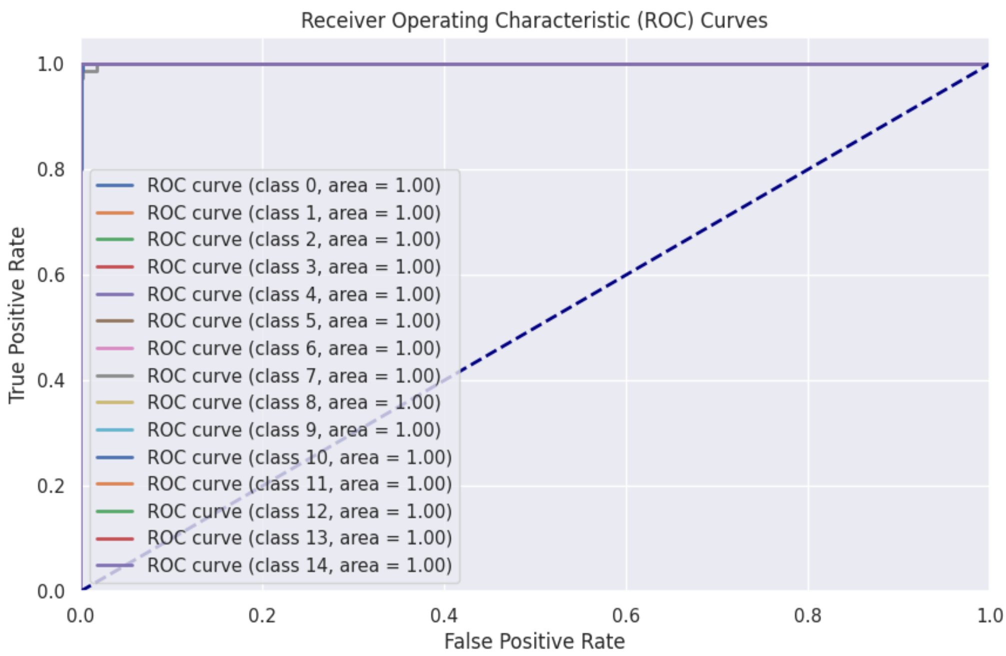

Figure 5.16: ROC Curve of EfficientNet-B4

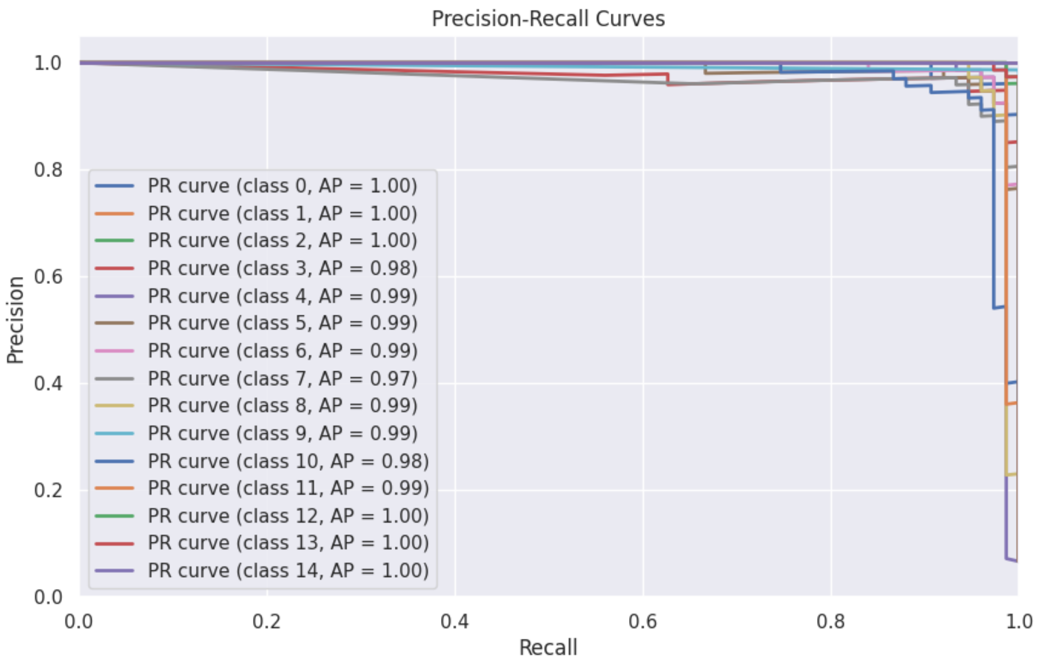

Figure 5.17: PR Curve of Basic CNN

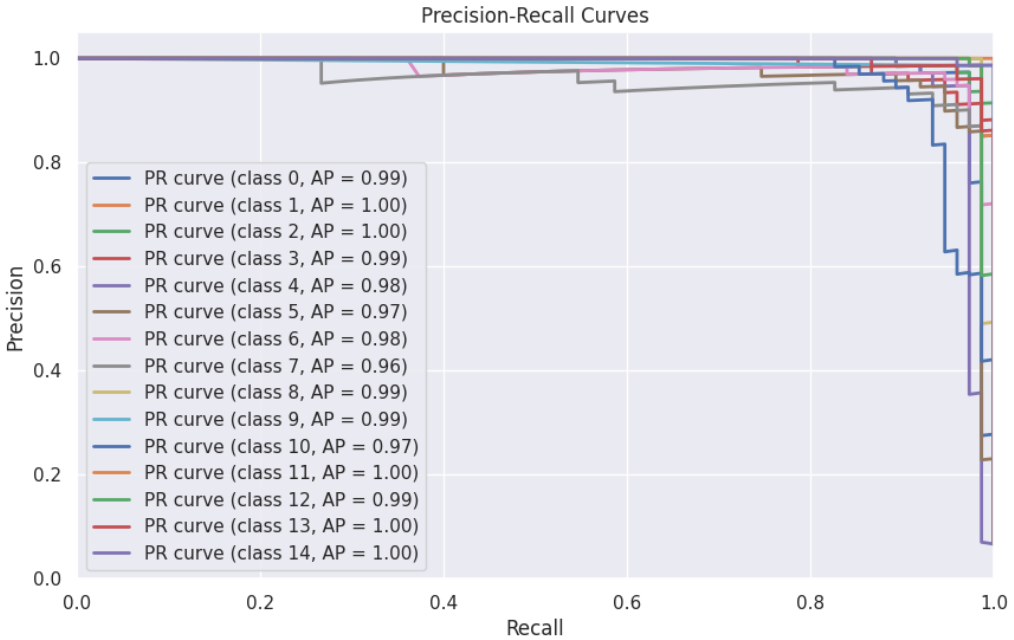

Figure 5.18: PR Curve of Modified AlexNet

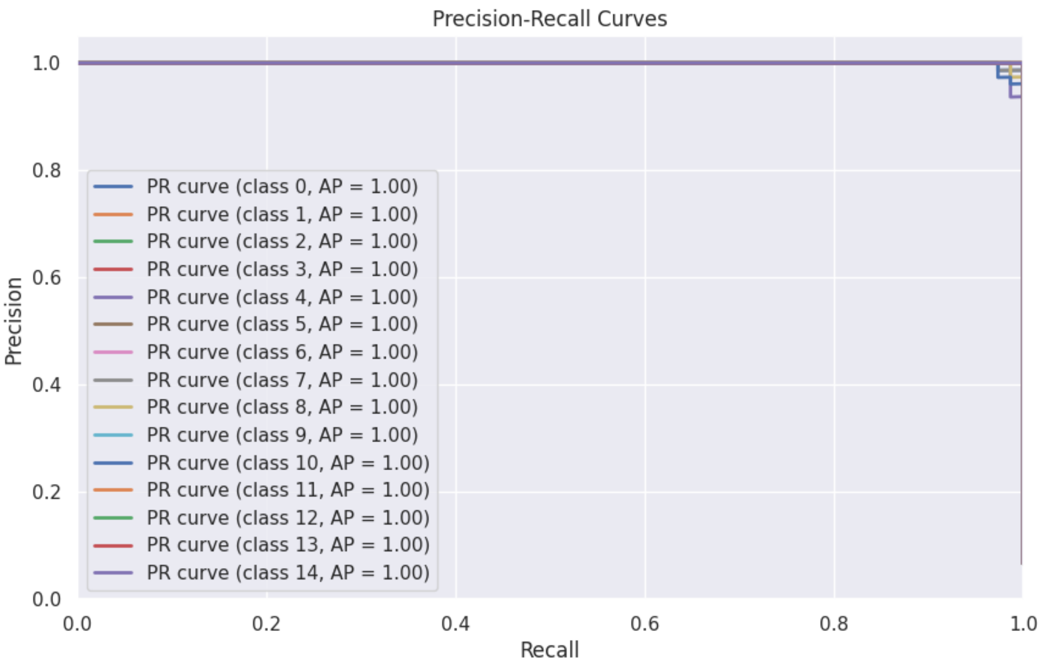

Figure 5.19: PR Curve of EfficientNet-B0

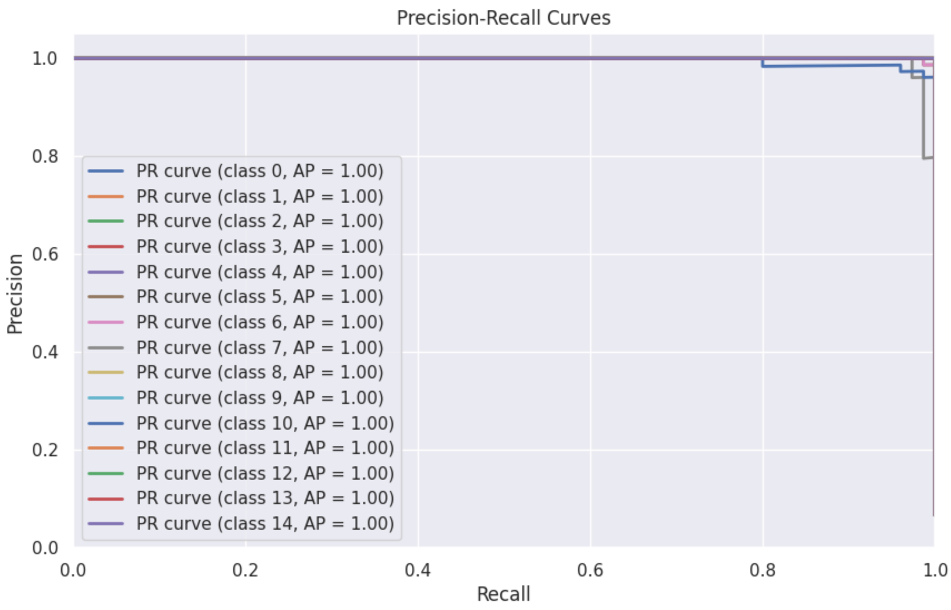

Figure 5.20: PR Curve of EfficientNet-B4

5.3.2 Comparative Analysis of Models

Within the field of agricultural disease diagnoses and management, a thorough examination of several convolutional neural network (CNN) models reveals insights into their effectiveness and efficiency. The models demonstrate remarkable accuracy, with the EfficientNet B4 model particularly noteworthy for achieving a flawless accuracy score of 100%. The Modified AlexNet and EfficientNet B0 models closely trail after, with accuracies of 99.6%. The Basic CNN model achieves an impressive accuracy of 99.5% shown in figure 5.21. Regarding loss measures, all models exhibit modest loss values, suggesting successful learning and resilient training. The Basic CNN model exhibits consistent precision and recall scores across several illness classes, leading to an overall F1-score of 0.968 when evaluating precision-recall and F1-score measures. Similarly, the Modified AlexNet and EfficientNet B0 models provide good precision-recall scores, but there are modest variances across different classes. Nevertheless, the EfficientNet B4 model outperforms all other models by obtaining flawless precision-recall scores and an immaculate F1-score of 1.0. By comparing the classification reports, it may have a clearer understanding of how well the models performed in terms of accuracy, recall, and F1-score metrics shown in figure 5.22. The Basic CNN model has balanced precision and recall scores, but the Modified AlexNet and EfficientNet B0 models provide improved accuracy and small enhancements in precision-recall metrics. However, the EfficientNet B4 model stands out as the most advanced solution, with unmatched accuracy, impeccable precision-recall scores, and immaculate categorization for all illness categories. Examining confusion matrices provides more understanding about the models' classification accuracy. The Basic CNN model demonstrates misclassifications, as shown by the non-diagonal values in the confusion matrix. In contrast, both the Modified AlexNet and EfficientNet B0 models exhibit enhanced classification accuracy, accompanied by a reduced number of misclassifications. The EfficientNet B4 model demonstrates outstanding performance by attaining flawless predictions for all classes, with no misclassifications reported in the confusion matrix. The EfficientNet B4 model is the leading solution in the area of agricultural disease diagnostics. It has revolutionized the sector by achieving exceptional accuracy, ideal precision-recall scores, and immaculate classification performance. The EfficientNet B4 model, via the utilization of sophisticated CNN architectures and state-of-the-art approaches like neural architecture search and compound scaling, establishes a higher benchmark for the categorization of plant diseases. This advancement also opens up possibilities for the development of more precise and dependable strategies for managing these diseases. The exceptional performance of deep learning highlights its potential to revolutionize agriculture and effectively tackle important concerns related to food security and crop protection.

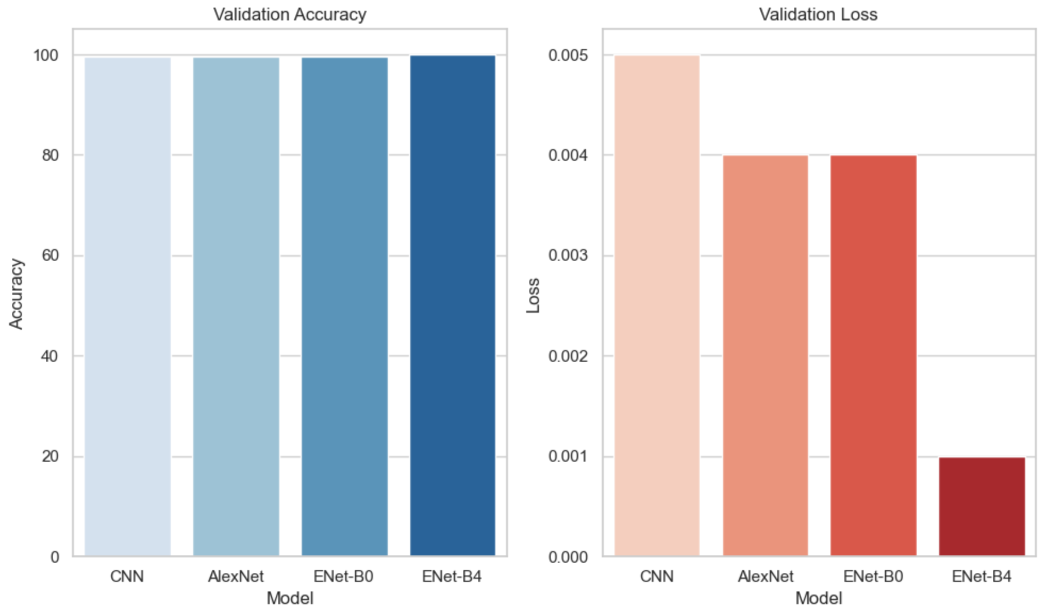

Figure 5.21: Comparison of Accuracy and Loss among All models

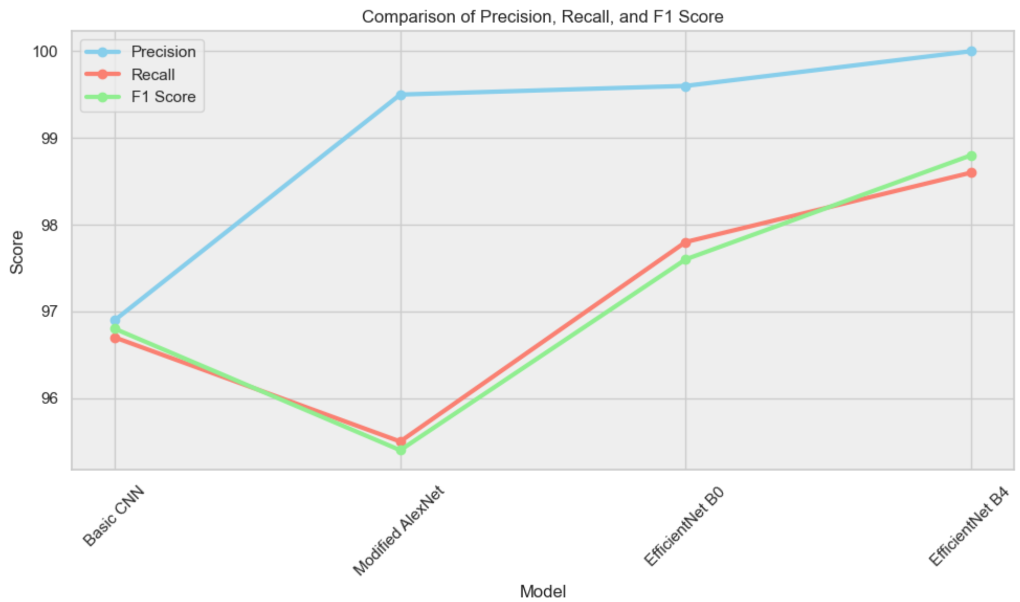

Figure 5.22: Comparison of Precision, Recall and F1 Score among All models.

5.3.3 Discussion

This thesis showcases a thorough investigation comparing the effectiveness of four different models for classifying plant diseases: Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. The main goal was to determine the optimal model for reliably detecting and categorizing different forms of plant illnesses, which is essential for early identification and efficient disease control in agriculture. The inquiry progressed via a comprehensive analysis, which included evaluating the correctness of the model, measuring loss, precision, recall, F1 score, and examining confusion matrices. Accuracy and loss are essential variables for assessing the overall performance and efficacy of machine learning models.The analysis found that EfficientNet B4 outperformed all other models in terms of accuracy, attaining a flawless score of 100%. The model's remarkable precision highlights its capacity to accurately categorize plants with diseases across many classifications. In addition, EfficientNet B4 exhibited the lowest loss compared to the other models, indicating its better ability to reduce prediction mistakes and improve predictive accuracy. EfficientNet B4 had exceptional accuracy and loss performance. It is important to note that all models achieved respectable accuracies ranging from 96.9% to 100%, indicating their effectiveness in classifying plant diseases. Precision, recall, and F1 score provide more profound understanding of a model's classification aptitude, specifically its capacity to accurately recognize positive instances while limiting both false positives and false negatives. EfficientNet B4 consistently demonstrated superior performance compared to other models, achieving the greatest accuracy, recall, and F1 score in this research. The model has extraordinary accuracy in correctly recognizing positive occurrences, strong ability to capture all positive examples, and an ideal balance between accuracy and completeness, as shown by the F1 score. The results highlight the effectiveness of EfficientNet B4 in precisely detecting and categorizing plant diseases, which is essential for making well-informed decisions in agricultural activities. An essential component of model assessment is the examination of confusion matrices, which provide a comprehensive breakdown of a model's classification accuracy in several illness categories. The confusion matrix of EfficientNet B4 demonstrated a reduced number of misclassifications and an increased number of right predictions in comparison to other models, thereby emphasizing its better accuracy in classification and its resilience. In contrast, whereas Basic CNN, Modified AlexNet, and EfficientNet B0 models had decent classification performance, they showed greater rates of misclassifications and reduced accuracy in certain illness categories. This implies that while these models may provide feasible alternatives, they may not achieve the same level of overall performance as EfficientNet B4 in illness classification tasks. The comparison study highlights EfficientNet B4 as the most advanced model for classifying plant diseases. It provides exceptional accuracy, little loss, and higher precision, recall, and F1 score compared to other models. The outstanding effectiveness of this technology may be credited to its sophisticated structure and effective usage of computer resources, making it a very attractive option for practical use in agriculture and plant pathology. However, more study and validation using a variety of datasets and environmental variables are necessary to confirm its reliability and applicability in real-world situations.

5.5 Summary

Chapter 5 of the thesis presents a thorough comparative examination of four distinct convolutional neural network (CNN) models—Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4—for the purpose of classifying plant diseases. The review aimed to measure the accuracy, loss, precision, recall, and F1 score parameters for different illness categories. The EfficientNet B4 model outperformed previous models, obtaining perfect accuracy and displaying higher precision-recall metrics. The chapter emphasises the efficacy of EfficientNet B4 in precisely classifying plant diseases, highlighting its potential to improve agricultural disease control and efforts to ensure food security.

CHAPTER 6
Impact on Society

6.1 Impact on Life

The findings from the research on deep learning models for plant disease classification have profound implications for society, particularly in the realm of agriculture and food security. By accurately identifying and classifying plant diseases through advanced artificial intelligence techniques, this research has the potential to revolutionize agricultural practices, enhance crop management strategies, and ultimately contribute to global food security. One of the most significant societal impacts of this research lies in its potential to mitigate crop losses caused by plant diseases. Agricultural productivity is often hindered by the prevalence of plant diseases, leading to substantial economic losses for farmers and food shortages for communities. The implementation of robust and accurate diagnostic tools based on deep learning models can empower farmers to identify and address plant diseases at an early stage, thereby minimizing crop losses and safeguarding food production. Furthermore, the accessibility and affordability of these advanced diagnostic solutions can extend their benefits to smallholder farmers and underserved agricultural communities. In many parts of the world, farmers lack access to timely and reliable disease diagnosis, leading to significant yield losses and economic hardship. By democratizing access to state-of-the-art plant disease classification technologies, this research has the potential to level the playing field for farmers, irrespective of their geographical location or socioeconomic status. Moreover, the integration of deep learning models into agricultural practices can promote sustainable farming methods and reduce reliance on chemical pesticides.

Traditional methods of disease management often involve indiscriminate pesticide use, which can have detrimental effects on human health, biodiversity, and the environment. By enabling more targeted and precise disease management strategies, these advanced diagnostic tools can help farmers adopt environmentally friendly practices, ultimately promoting agricultural sustainability and ecological balance.Beyond agricultural productivity, the societal impact of this research extends to broader environmental conservation efforts. Healthy plants play a crucial role in ecosystem functioning, carbon sequestration, and climate regulation. By facilitating the early detection and management of plant diseases, deep learning models contribute to preserving biodiversity, maintaining ecosystem resilience, and mitigating the impacts of climate change on agricultural systems. The research on deep learning models for plant disease classification holds immense potential to transform agricultural practices, enhance food security, and promote environmental sustainability. By leveraging cutting-edge technologies, this research not only advances the scientific understanding of plant pathology but also addresses pressing societal challenges related to agriculture, health, and the environment. Ultimately, the widespread adoption of these advanced diagnostic tools has the capacity to improve the livelihoods of farmers, ensure food security for communities, and promote sustainable development on a global scale.

6.2 Impact on Society & Environment

The study on deep learning models for plant disease categorization has important implications for environmental conservation and sustainability. These modern technologies may help improve disease management in agriculture by making it more accurate and focused. As a result, they have the potential to reduce environmental degradation, boost ecosystem health, and contribute to larger conservation efforts. An important environmental advantage of this study is its ability to decrease dependence on chemical pesticides. Conventional approaches to disease control often rely on the widespread use of pesticides, which may negatively impact soil fertility, water purity, and unintended species. Deep learning models enable farmers to identify and diagnose diseases at an early stage, which in turn allows them to use integrated pest management (IPM) strategies. IPM focuses on using non-chemical approaches including biological control, crop rotation, and cultural practices.

The use of more sustainable pest control methods reduces the negative impact of agriculture on the environment, ensuring the preservation of ecosystem integrity and biodiversity. Furthermore, the use of sophisticated disease detection technology may aid in diminishing agricultural runoff and the pollution of aquatic bodies by pesticides. The overutilization of pesticides in current agricultural methods may result in water contamination, which presents hazards to aquatic ecosystems, animals, and human well-being. Deep learning models assist farmers in implementing precise and careful pesticide applications, therefore reducing environmental hazards and safeguarding water quality and aquatic ecosystems. Consequently, this aids in the preservation of freshwater ecosystems, fosters the variety of aquatic life, and guarantees the accessibility of uncontaminated water for diverse purposes. In addition, deep learning models indirectly help reduce the need for converting land and deforestation by limiting agricultural losses caused by plant diseases. The development of agricultural land in several places worsens habitat degradation, loss of biodiversity, and fragmentation of natural ecosystems as a means to compensate for yield losses. These modern technologies enhance crop health and production, hence optimizing agricultural output on current farmland. This reduces the need for further land conversion and helps protect natural ecosystems. This conservation advantage also helps reduce greenhouse gas emissions linked to land-use change and deforestation, therefore aiding in climate change mitigation efforts.

Implementing disease control measures based on deep learning techniques enhances soil health and fertility by decreasing the need for extensive plowing and chemical inputs. Implementing sustainable farming techniques such as minimizing tillage and utilizing cover crops may strengthen soil structure, increase water retention, and improve nutrient cycling. These activities ultimately result in enhanced soil health and resilience. Deep learning models aid in the adoption of regenerative agriculture techniques, promoting soil conservation, carbon sequestration, and long-term sustainability in agricultural landscapes. The investigation into deep learning models for the categorization of plant diseases has encouraging prospects for improving environmental sustainability in the field of agriculture. These innovative technologies help to the development of resilient and ecologically responsible agricultural systems by encouraging sustainable pest management techniques, lowering pesticide usage, protecting water quality, conserving natural habitats, and supporting soil health. In conclusion, the extensive use of disease control methods based on deep learning has the capacity to promote a harmonic equilibrium between agricultural output and environmental preservation, guaranteeing the sustainability of food production systems in a shifting climate.

6.3 Ethical Aspects

The progress of deep learning models for plant disease categorization raises ethical concerns that need thorough scrutiny and proactive steps to guarantee appropriate implementation and use of these technologies. Like other technical advancements, ethical concerns emerge throughout different phases of development, implementation, and use, requiring a comprehensive approach to ethical decision-making and governance. Data privacy and permission are key ethical issues. The construction and training of deep learning models depend on extensive datasets, which include pictures of plant diseases gathered from many sources. It is crucial to adhere to the rules of data privacy and get informed permission from people or organizations that provide data for research purposes. Transparent data collecting procedures, explicit permission processes, and compliance with data protection rules are essential for protecting the privacy rights of individuals and ensuring confidence in research endeavors. Algorithmic bias and fairness are additional ethical considerations to be taken into account. Deep learning models are vulnerable to biases that are inherent in the training data, which might result in unjust or discriminating results, especially in healthcare decision-making. Researchers must confront biases present in training datasets, alleviate algorithmic biases via meticulous validation and testing, and guarantee fairness and equality in model predictions across various demographic groups. Ethical issues also include the conscientious reporting and interpretation of research findings, refraining from sensationalism or exaggeration of outcomes that may foster impractical expectations or improper use of technology.

Moreover, the proper use of deep learning models requires careful thought on accountability and transparency. The stakeholders engaged in the creation, validation, and deployment of these technologies have ethical obligations to guarantee the dependability, security, and honesty of model outcomes. Clear and comprehensive documentation of the model's structure, training methods, and validation procedures promotes peer review, the capacity to reproduce results, and responsibility in research activities. Furthermore, effectively conveying the constraints, uncertainties, and possible hazards associated with a model aid in setting realistic expectations and promoting well-informed decision-making among users and stakeholders.

Ethical issues also include the wider social effects and consequences of using technology. Plant disease categorization using deep learning models has the capacity to significantly impact current agricultural practices, livelihoods, and socio-economic dynamics within farming communities. To guarantee the fair distribution of benefits, it is crucial for researchers and policymakers to include stakeholders such as farmers, agricultural extension workers, and policymakers in ethical discussions and participatory decision-making processes throughout the development of technology. Implementing proactive strategies to tackle the possibility of job displacement, skill deficiencies, and socio-economic disparities resulting from the use of technology is crucial for promoting ethical and socially conscious advancements in agriculture. Essentially, ethical issues are crucial in the process of creating, implementing, and using deep learning models for the categorization of plant diseases. Adhering to the values of data privacy, fairness, openness, and accountability is crucial for doing research responsibly and fostering trust, equality, and societal well-being. To effectively tackle ethical concerns in agriculture, researchers may adopt a proactive approach and include stakeholders in ethical discussions. This way, they can use sophisticated technology to solve urgent issues while maintaining ethical norms and values. The sustainability strategy for deep learning models in plant disease categorization incorporates many essential components to guarantee long-term durability and inclusiveness. The process starts by completing comprehensive environmental impact evaluations to comprehend and alleviate probable environmental consequences. Advocating for the use of sustainable energy sources and energy-efficient computer systems aids in the reduction of resource utilization and the mitigation of carbon emissions.

Data privacy and security safeguards protect confidential data, guaranteeing adherence to legal and ethical guidelines. Capacity building and information sharing efforts enable agricultural stakeholders to efficiently use technology, while socio-economic impact studies guide fair decision-making and policy development. Stakeholder involvement and governance frameworks enhance transparency, accountability, and social legitimacy, hence facilitating cooperation and promoting inclusive development in agricultural communities. The sustainability plan aims to use technology to solve global concerns and promote sustainability and resilience in agricultural ecosystems by combining environmental stewardship, social responsibility, and economic viability.

6.4 Sustainability Plan

The sustainability strategy for deep learning models in plant disease categorization incorporates many essential components to guarantee long-term durability and inclusiveness. The process starts by completing comprehensive environmental impact evaluations to comprehend and alleviate probable environmental consequences. Advocating for the use of renewable energy sources and energy-efficient computer systems aids in the reduction of resource consumption and carbon emissions. Data privacy and security safeguards protect sensitive information, guaranteeing adherence to legal and ethical guidelines. Capacity building and information sharing efforts enable agricultural stakeholders to efficiently use technology, while socio-economic impact studies guide fair decision-making and policy development. Stakeholder involvement and governance frameworks enhance transparency, accountability, and social legitimacy, hence facilitating cooperation and promoting inclusive development in agricultural communities. The sustainability plan aims to combine environmental stewardship, social responsibility, and economic viability in order to use technology's capabilities for tackling global issues and fostering sustainability and resilience.

6.5 Summary

The study investigates the significant socioeconomic, environmental, ethical, and sustainability effects of using deep learning models for the categorization of plant diseases in agriculture. This statement emphasises the potential of sophisticated technology to transform agricultural practices via improved crop management, reduced crop losses caused by diseases, and the promotion of sustainable farming techniques. The chapter highlights the capacity of deep learning models to democratise the availability of diagnostic tools, enhance food security, and alleviate environmental degradation via the reduction of pesticide use and the promotion of ecosystem health. The discussion also encompasses ethical concerns such as safeguarding data privacy, addressing algorithmic prejudice, and ensuring equitable distribution of benefits.

Additionally, a holistic sustainability strategy is advocated, which promotes environmental stewardship, social responsibility, and inclusive growth in agricultural communities.

CHAPTER 7
Summary, Conclusion,
Recommendation and Implication for Future Research

7.1 Conclusions

This research investigated the effectiveness of deep learning models in classifying plant diseases. Four different architectures were compared: Basic CNN, Modified AlexNet, EfficientNet B0, and EfficientNet B4. By conducting rigorous training and assessment, the performance of each model was tested based on accuracy, loss metrics, precision, recall, and F1-score. The results indicated that all models performed well, but EfficientNet B4 stood out as the best performer, showing exceptional accuracy and precision in several illness categories. The hierarchical structure and sophisticated convolutional layers of the model allowed it to accurately detect complex patterns in photos of plant diseases. Furthermore, the research emphasized the social implications of these models, underscoring their capacity to transform agricultural disease diagnosis, expedite healthcare procedures, and enhance overall public health results. Nevertheless, it is crucial to meticulously include ethical concerns, environmental ramifications, and sustainability measures into the implementation of these models to guarantee responsible and fair consumption. In summary, this research highlights the significant impact that deep learning may have on agriculture and healthcare, by enabling the development of disease control systems that are more precise, effective, and environmentally friendly.

7.2 Further Suggested Works

Although this work offers interesting insights into the use of deep learning models for classifying plant diseases, it is important to highlight several limitations. Firstly, the assessment was largely done on a particular dataset, which may restrict the generalizability of the results to other datasets or real-world circumstances. Furthermore, the efficacy of the models might differ based on variables such as picture resolution, illumination circumstances, and the existence of extraneous components in the surroundings. To overcome these constraints, future study should undertake comprehensive assessments on a wide range of datasets and investigate the robustness of the findings under various environmental situations. In addition, the research largely concentrated on classifying images and did not investigate other types of data, such as spectral or temporal data, which might provide additional information for detecting diseases. Subsequent research might explore the incorporation of several data sources from different modes to improve the accuracy of the model and expand the range of illness detection capabilities. Furthermore, while attempts were made to enhance the performance of the model by adjusting hyperparameters and selecting the architecture, there is still potential for further optimization and improvement. Subsequent research endeavors may investigate sophisticated methodologies for optimizing models, using ensemble learning approaches, and enhancing model interpretability. These efforts aim to improve performance and enable the deployment of models in real-world scenarios. This research did not thoroughly discuss the ethical concerns related to data privacy, algorithmic bias, and model interpretability.

Further investigation should focus on these ethical dimensions to guarantee the acceptable development and use of AI technologies in the fields of agriculture and healthcare. Future research must prioritize the environmental sustainability of deep learning models, specifically in relation to energy usage and carbon footprint. Exploring methods to enhance model efficiency, minimize resource use, and encourage sustainable behaviors might aid in mitigating environmental effects and guaranteeing the enduring feasibility of AI-driven solutions in agriculture and healthcare. The results of this work emphasize several opportunities for more investigation and development in the domain of plant disease categorization using deep learning techniques. Future research has the potential to enhance the comprehension of AI technologies in agriculture and healthcare by addressing the limitations identified and building upon the insights obtained from this study. This, in turn, can result in more efficient disease management strategies, better public health outcomes, and the adoption of sustainable agricultural practices.

Examining the transferability of pre-trained models across different crops and illnesses has the potential to create flexible diagnostic tools that may be used in numerous agricultural fields. Furthermore, the incorporation of diverse data sources, such as spectral or temporal information, has the potential to improve the resilience and precision of illness detection systems. Ensuring justice, openness, and accountability in agricultural processes requires rigorous evaluations and rules for appropriate AI implementation, with ethical issues being of utmost importance. In addition, enhancing model efficiency via the exploration of methods to decrease the computing resources needed for training and inference may enhance scalability and promote environmental sustainability. Field validation and deployment are essential stages that need cooperation with stakeholders to verify the model's performance in actual agricultural environments and promote its acceptance. By focusing on these study topics, future studies may make substantial contributions to the advancement of AI-driven solutions for the diagnosis and treatment of plant diseases. This has wider implications for agriculture, public health, and environmental conservation.

7.3 Limitations/Conflict Interests

After conducting a thorough investigation and presenting the results in this thesis, we have noticed many limitations and possible conflicts of interest. These findings should be taken into account for future research and application. An important constraint is associated with the capacity to apply the results to a wider population. Although the deep learning models, specifically EfficientNet B4, showed strong performance in accurately categorising plant diseases in the specific datasets and experimental conditions used in this study, their performance may differ when applied to different datasets with varying characteristics like image quality, resolution, and environmental factors. Future research should focus on validating these models using a wide range of datasets and real-world agricultural settings in order to improve their dependability and practicality. One further constraint is to the moral considerations and possible prejudices inherent in the creation and use of AI systems in agriculture. It is crucial to tackle issues related to data privacy, algorithmic fairness, and openness in decision-making processes in order to guarantee appropriate and fair utilisation of these technologies. In order to address these issues and establish confidence among end-users and impacted communities, future research should include rigorous ethical frameworks and stakeholder engagement methodologies. Moreover, the real implementation of training and deploying complex deep learning models such as EfficientNet B4 requires significant computing resources and infrastructure. These models often need significant processing power and storage, which might be a barrier for small-scale farmers or resource-limited environments. Subsequent investigations should focus on examining techniques to enhance the effectiveness of the model, decrease the computational expenses, and create versions that are both easily accessible and adaptable to other agricultural settings.

In addition, while this thesis mainly focused on the categorization of plant diseases using images, further investigation might examine the incorporation of other data modalities such as spectral data, temporal data, or sensor data from Internet of Things (IoT) devices. By integrating these many data sources, the accuracy and dependability of disease detection systems may be improved, resulting in a more thorough understanding of crop health and disease patterns across time. It is important to openly declare any conflicts of interest that may arise from funding sources, industrial relationships, or commercial implementations of the study results. Transparent disclosure of any conflicts guarantees the honesty and fairness of the study results, promoting reliability and confidence within the scientific community and interested parties. To advance the field of AI-driven plant disease classification and ensure responsible deployment, it is crucial to address these limitations and conflicts of interest. This will maximise the potential of AI to positively contribute to agricultural sustainability, food security, and environmental conservation.Transparency of models, guarantee fair AI implementations, and carefully conform to ethical guidelines and data privacy laws.

References

[1] J. S. Dias, "Nutritional quality and health benefits of plant diseases: a review", *Food & Nutrition Science, vol. 3, pp. 1354–1374, 2012.*

[2] J. Keatinge, R.-Y. Yang, J. A. Hughes, W. Easdown, and R. Holmer, "The importance of plant diseases in ensuring both food and nutritional security in attainment of the Millennium Development Goals", *Food Security, vol. 3, pp. 491–501, 2011.*

[3] M. Pitrat and J. M. Audergon, "Evolution of diversity of plant diseases and plant diseases crops", *Acta Horticulturae, vol. 1099, pp. 567–576, 2015.*

[4] G. E. Ilbaum, Plant disease History, Nomenclature, and Classification. *Wallingford: CAB International, 2015, pp. 1–15.*

[5] T. Radovich, "Biology and Classification of Plant diseases", *in Plant Disease, 1st ed., 2018. doi: 10.1002/9781119098935.ch1.*

[6] S. C., N. Manasa, V. Sharma, and N. K. A. A., "Plant disease Classification Using You Only Look Once Algorithm"," *in 2019 International Conference on Cutting-edge Technologies in Engineering (ICon-CuTE), Uttar Pradesh, India, 2019, pp. 101-107. doi: 10.1109/ICon-CuTE47290.2019.8991457.*

[7] L. E. Graham, J. M. Graham, and L. W. Wilcox, Plant Biology, 2nd ed. Upper Saddle River, NJ: Pearson/Prentice-Hall, 2006, p. 704.

[8] Z. Li, F. Li, L. Zhu, and J. Yue, "Plant disease Recognition and Classification Based on Improved VGG Deep Learning Network Model", *International Journal of Computational Intelligence Systems, vol. 13, pp. 559, 2020. doi: 10.2991/ijcis.d.200425.001.*

[9] H. Singh, R. Singh, P. Goel, A. Singh, and N. Sharma, "Automatic Framework for Plant disease Classification using Transfer-Learning", *International Journal of Electrical and Electronics Research*, vol. 10, pp. 405-410, 2022. doi: 10.37391/ijeer.100257.

[10] M. I. Ahmed, S. Mamun, and A. Asif, "DCNN-Based Plant disease Image Classification Using Transfer Learning: A Comparative Study", *in 2021 IEEE International Conference on Computing, Communication, and Security (ICCCSP), 2021, pp. 235-243. doi: 10.1109/ICCCSP52374.2021.9465499.*

[11] G. Zeng, "Plant disease and plant diseases classification system using image saliency and convolutional neural network", in 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC), *Chongqing, China, 2017, pp. 613-617. doi: 10.1109/ITOEC.2017.8122370.*

[12] N.-Q. Pham, T.-S. Nguyen, J. Niehues, et al., "Very deep self-attention networks for end-to-end speech recognition", arXiv preprint arXiv: 1904.13377, 2019.

[13] L. Zhu, Z. Li, C. Li, et al., "High-performance plant disease classification from images based on AlexNet deep learning model", *International Journal of Agricultural and Biological Engineering, vol. 11, pp. 217–223, 2018.*

[14] S. Ciptohadijoyo, W. S. Litananda, M. Rivai, et al., "Electronic nose based on partition column integrated with gas sensor for plant disease identification and classification", *Computers and Electronics in Agriculture, vol. 121, pp. 429–435, 2016.*

[15] K. Hameed, D. Chai, and A. Rassau, "A comprehensive review of plant disease and plant disease classification techniques", *Image and Vision Computing*, vol. 80, 2018. doi: 10.1016/j.imavis.2018.09.016.

[16] J. M. Alston and P. G. Pardey, "Agriculture in the Global Economy", *Journal of Economic Perspectives, vol. 28, pp. 121–146, 2014.*

[17] "Contribution of Agriculture Sector Towards GDP Agriculture Has Been the Bright Spot in the Economy despite COVID-19," [Online]. Available: https://www.pib.gov.in/indexd.aspx. [Accessed: 29-May-2024].

[18] L. Li, S. Zhang, and B. Wang, "Plant Disease Detection and Classification by Deep Learning—A Review", *IEEE Access*, vol. 9, pp. 56683–56698, 2021.

[19] A. Sharma, T. Ahmad, and S. Upadhyay, "Deep Learning for Plant Diseases: Detection and Saliency Map Visualisation", *in 2019 2nd International Conference on Inventive Research in Computing Applications (ICIRCA), 2019, pp. 693-697.*

[20] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics", arXiv preprint arXiv:1511.08060, 2015.

[21] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition", *Computers and Electronics in Agriculture, vol. 176, pp. 105568, 2020.*

[22] R. Mehra, R. Kaundal, and D. Gupta, "A Comprehensive Review on Automated Plant Disease Recognition and Classification Techniques", *IEEE Access, vol. 9, pp. 31725-31745, 2021.*

[23] T. T. Santos, C. D. Cruz, E. Pedrozo, and B. R. Telles, "A Comparative Study of Machine Learning Techniques for Plant Disease Identification", *in 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), 2018, pp. 1057-1064.*

[24] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection", *Frontiers in Plant Science, vol. 7, pp. 1419, 2016.*

[25] M. T. Islam, A. Asraf, and M. A. Moni, "Plant Disease Detection and Classification by Deep Learning", *in Deep Learning Techniques for Biomedical and Health Informatics, Springer, Cham, 2019, pp. 273-287.*

[26] A. Haghighattalab, J. G. González-Durán, A. W. Schumann, M. C. Whalen, and D. C. Iindorf, "Machine Learning for Agricultural Field Phenotyping: *A Review", ISPRS International Journal of Geo-Information, vol. 9, pp. 498, 2020.*

[27] N. Ullah, J. A. Khan, S. Almakdi, M. S. Alshehri, M. Al Qathrady, N. El-Rashidy, S. El-Sappagh, and F. Ali, "An effective approach for plant leaf diseases classification based on a novel DeepPlantNet deep learning model", *Frontiers in Plant Science, vol. 14, pp. 1212747, 2023. doi: 10.3389/fpls.2023.1212747.*

[28] A. K. P. and J. Anitha, "Plant disease classification using deep learning", in 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp. 407-411. doi: 10.1109/ICSPC51351.2021.9451696.

[29] M. Sibiya and M. Sumbwanyambe, "A Computational Procedure for the Recognition and

Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks", *AgriEngineering, vol. 1, pp. 119-131, 2019. doi: 10.3390/agriengineering1010009.*

[30] K. Zhang, Q. Wu, A. Liu, and X. Meng, "Can Deep Learning Identify Tomato Leaf Disease?", *Advances in Multimedia*, vol. 2018, pp. 10, 2018. doi: 10.1155/2018/6710865.

[31] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection", *Frontier in Plant Science*, vol. 8, pp. 1852, 2017.

[32] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: *A comprehensive tutorial with selected use cases", Journal of Pathology Informatics, vol. 2016.*

[33] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence", *Nature Medicine, vol. 25, no. 1, 2019.*

[34] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," *in 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), 2017, pp. 721-724. doi: 10.1109/ICBDA.2017.8078730.*

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title: Plant Disease Classification Using Deep Learning Approaches

Student ID: 202-15-14416

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify Plant disease **detection** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goal for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**

CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**
CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, and CO3	My project requires data collection from the field and I have to know the design of a machine learning-based model which covers K3 and K4	**Page no: 13-22** **Section:** [3.1, 3.2]

				Depth of analysis needed to build a model from scratch.	**Page no: 22-29** **Section:** [3.2]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2	In the project, the strength of the Model and the datasets are directly related to the capacity of system	**Page no: 22-28** **Section:** [3.2]
3.	**EP3: Depth of analysis required**	Yes	CO2	The project requires knowledge of Artificial Intelligence, Machine Learning, Deep learning, convolution neural network, CNN Architecture, Transfer learning.	**Page no: 11-32** **Section:** [3.1, 3.2, 3.3, 3.4, 3.5]
4.	**EP4: Familiarity of Issues**	Yes	CO3	Noisy dataset, GPU Resource	**Page no: 11-32** [3.1, 3.2, 3.3, 3.4, 3.5]
5.	-	Yes	CO4	I have prepared a budget to evaluate and estimate the cost required for my FYDP.	**Page no: 26** [3.4]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	My project utilizes diverse resources such as high-performance computing infrastructure, GPUs, deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in pneumonia detection through transfer learning and deep CNNs.	**Page no:** [15-27] **Section:** [3.2]
2.	**EA4: Consequences for society and the environment**	Yes	CO11	This project contributes to society by improving agriculture through advanced Plant disease detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.	**Page no:** [38+6-64] **Section:** [5.1, 5.2, 5.4]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [26] **Section:** [3.4]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing patient privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of advanced agriculture technologies which comply with **CO9**.	**Page no:** [35] **Section:** [5.3]
3	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [24-27, 28-32] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2]

PLANT DISEASE CLASSIFICATION USING DEEP LEARNING APPROACHES

ORIGINALITY REPORT

21%	16%	13%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	5%
2	www.mdpi.com Internet Source	2%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	Theodore J. K. Radovich. "Biology and Classification of Vegetables", Wiley, 2018 Publication	1%
5	www.atlantis-press.com Internet Source	1%
6	www.frontiersin.org Internet Source	1%

7	"Advanced Research in Technologies, Information, Innovation and Sustainability", Springer Science and Business Media LLC, 2021 Publication	1%
8	ebin.pub Internet Source	1%
9	R Aparna, Shilpa Chaudhari, Kapsha Roshan Singh, Fayaz Ahmed Mohammad, Manish Padmannavar, Varra Charishma. "Video Analytics-based Ingredient Detection for Smart Chef using YOLOv3 and ImageAI", 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), 2021 Publication	<1%
10	digital.library.unt.edu Internet Source	<1%
11	Submitted to Coventry University Student Paper	<1%
12	www.irejournals.com Internet Source	<1%

13	Putta Durga, Dhiren Gangishetty, Deepthi Godavarthi. "Convolutional Neural Network (CNN) Based Identification of Crop Diseases Using ResNet-50", 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), 2023 Publication	<1%
14	digibug.ugr.es Internet Source	<1%