Identification of Potato leaf disease using lightweight transformer architecture and mobile app-based detection

By

MD TAMIM HOSSAIN

ID: 213-15-4605

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised by

Mr. Shahadat Hossain
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

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

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 17, 2025

APPROVAL

This Project titled "**Identification of Potato leaf disease using lightweight transformer architecture and mobile app-based detection**," submitted by MD TAMIM HOSSAIN, ID No: **213-15-4605** 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 **17 September, 2025**

BOARD OF EXAMINERS

Dr. Bimal Chandra Das (BCD) **Chairman**
Professor and Dean (in-charge)
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Shah Md Tanvir Siddique (SMTS) **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Hasanuzzaman Dipu (MHD) **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Nazibur Rahman **External Examiner**
Technical Lead · Database Administrator
Telenor · Grameen Phone Account

DECLARATION

We hereby declare that this project has been done by us under the supervision of Mr. Shahadat Hossain, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

09/09/25

Mr. Shahadat Hossain
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by:

Shah Md Tanvir Siddique
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Submitted by:

Tamim

MD TAMIM HOSSAIN
Student ID: 213-15-4605
Department of Computer Science and EngineeringDaffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has as- sisted us in one way or another.

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

We are grateful and wish our profound indebtedness to **Mr. Shahadat Hossain**, **Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Deep Learning** to carry out this project. His endless pa- tience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

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

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

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

ABSTRACT

Agriculture faces significant challenges from plant diseases, with potato production particularly vulnerable to early and late blight, leading to substantial yield losses and food insecurity. Conventional diagnostic methods are often slow, costly, and inaccessible to resource-constrained farmers, underscoring the need for affordable, scalable solutions. This study proposes a lightweight Vision Transformer (ViT)–based framework integrated with a mobile application to enable real-time, on-field detection of potato leaf diseases. A curated dataset of 6,954 potato leaf images across three classes—Early Blight, Late Blight, and Healthy—was preprocessed using resizing, augmentation, and normalization to ensure robustness. The proposed MobileViT architecture, optimized for efficiency, was trained and benchmarked against state-of-the-art CNN models (EfficientNetV2S/M/L, ConvNeXtBase) under identical experimental settings. Results demonstrated that the lightweight ViT achieved superior performance, attaining 99.91% validation accuracy and 100% test accuracy with only 1.3M parameters and a 4.99 MB model size, significantly outperforming larger models in both accuracy and computational efficiency. Confusion matrix analyses confirmed flawless class-wise classification, while ROC-AUC scores of 1.00 validated its reliability in distinguishing visually similar disease symptoms. The trained model was deployed to a mobile application via TensorFlow Lite, enabling farmers to capture or upload leaf images for instant disease diagnosis offline, thus ensuring usability in low-connectivity regions. Beyond technical performance, the system offers economic and environmental benefits by reducing crop losses, minimizing pesticide misuse, and enhancing sustainable agricultural practices. This research bridges the gap between advanced deep learning architectures and their practical field deployment, presenting a scalable, cost-effective, and farmer-friendly tool for precision agriculture and contributing to global food security.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents the research problem, aims, and importance of the creation of a lightweight transformer-based model to detect the potato leaf disease. It emphasizes upon motivation, research questions, and scope of the study and also gives a structure of the thesis.

1.1 Introduction

The global food security highly depends on agriculture, and the potato is one of the most popular crops present in the world. Nevertheless, it is extremely susceptible to the common leaf illnesses like the late blight, the early blight, and other types of fungal or bacterial illnesses. Such diseases may result in the severe loss of yield, which impacts both the income of farmers and the chain of food provision. It is necessary to identify these diseases in time and correctly, then it is possible to take preventive or corrective measures. Conventional ways of detecting diseases are usually based on visual examination of educated perception or in the laboratory, and these processes are slow, expensive, and challenging to use in farmed areas [4][5][6]. Digital solutions, especially image-based diagnostics, have proven to be of high potential to accelerate, democratize, and make plant disease identification more accurate in the past couple of years [1].

With the development of computer vision and deep learning, the way agricultural issues have been approached changed. Convolutional Neural Networks (CNNs) are already employed in the eliquid studies precisely in the domain of plant disease classification [8], but lack the capacity to resolve long-range dependencies in images. Vision Transformers (ViTs) [10] are introduced initially to work on image recognition, which gives a portfolio of applying self-attention mechanisms to capture global connections of image patches, rendering them exceptionally suitable to the domain of fine-grained classification onsets. Other lightweight versions like MobileViT [9] and being hybridized between CNN and transformer [3][13][18] models has also made real-time and mobile-compatible disease detection a possibility. These architectures coupled with preprocessing such as resizing, Augmentation, and normalization have been performing robustly even when field conditions are difficult [17][19]. Also, app development has enabled an application to provide diagnosis to the smartphones of the farmers, facilitating time-saving decisions rendered in the field where AI is implemented [11][12].

These technological advancements combined with the needs of agriculture make up a potent possibility of pragmatic solutions. Applied to the task of detecting the potato leaf disease, the use of lightweight architectures of transformers, e.g. MobileViT, provides excellent accuracy at a low cost of computation, which makes it possible to deploy it on mobile devices with no performance drop [18][19]. This study enables addressing the accuracy problem and the gap in accessibility to this technology by the farmers since all the stages of image preprocessing, model training, evaluation, and mobile app development will be united in a single workflow, as follows in the research methodology. Through this, it will help in realizing sustainable agriculture as it will allow timely action, prevent misuse of chemicals, and prevent crop losses [3][13][24].

1.2 Motivation

The importance of potato production is connected with food security of the whole world and life of millions of farmers. Diseases like late blight and early blight are however a severe challenge which may result into a major loss of yield in case of non-early detection and care [4][5]. In contrast, the conventional means of identification (based on human operator observation or laboratory analysis) can be both costly and time-consuming, not to mention being unavailable to most of the farmers in rural or resource scarce locales [6]. Due to the rise of disease and climate change as well as the world trading economy, there is a pressing need to develop a disease detection system that is fast, accurate and affordable, and does not have the need to be monitored by an expert [1]. The solution to this need does not only save crops production, but also constrains the abuse of pesticides that may cause environmental and human damage.

Deep learning and computer vision research have in recent years appeared as the possible solution to address this challenge. The accuracy of vision transformer (ViT) architectures [10] and their lightweight models, like MobileViT [9][18] have shown that such models can achieve a high accuracy in plant leaf disease classification, and are lightweight to such an extent that they can be executed on mobile devices. Recently introduced, e.g., EfficientRMT-Net

[13], and PMVT [18] hybrid CNN-Transformer models make use of the advantages of convolutional and attention-based architectures and lead to robust detection regardless of the field conditions [17][19]. By adopting such architectures into a mobile software, the farmers would have a practical on-site diagnostic tool [11][12]. The impetus lying behind this thesis is synergistic combination of these technological breakthroughs with real world agricultural needs in order to develop an effective, precise, and farmer friendly system to detect potato leaf diseases that can be scaled up and applied to ensure sustainability in agricultural practices.

1.3 Objectives

The primary goal of this study is to develop and create an effective system of recognizing the diseases of the potatoes on the leaves with the help of a lightweight transformer-based architecture, as well as a mobile application to realize it in real-time. By combining high-tech deep learning-based algorithms with the practical mobile implementation, the aim of the proposed project will be presented in the fast, accurate, and intuitive and user-friendly solution among the farmers and agricultural workers. Its aims are the following ones:

- To Develop an optimized ViT or MobileViT-based model that will identify the potato leaf disease with a high accuracy rate using limited computing power; hence, can successfully be implemented to mobile phones or devices.
- To preprocess and augment potato leaf image datasets to achieve better model performance Use preprocessing steps which include resizing, normalization, augmentation, which ensures that the model would respond to other conditions of lighting, backgrounds and orientations to the leaf.
- To slot the trained model into a mobile application alongside real-time detection Utilize and deploy the trained model in an application that would allow farmers to take images of the leaves and get identifications of the diseases in real-time without the requirement of a high-end device or internet connection.
- To estimate model performance with standard metrics Test the model on the basis of accuracy, precision, recall and F1-score, and test its reliability under various environmental conditions and unseen data.

- To facilitate adoption and accessibility in agricultural practices To enable the ease of use, lightweight, and cost-effective system that offers early disease identification, optimal crop health, and greater agricultural efficiency.

1.4 Methodology

The intended research design uses a systematic method of creating an effective diabetes scan system of potato leaves through lightweight transformer models and the implementation of mobile apps. The first step entails data gathering wherein pictures of potato leaves are obtained online in publicly accessible repositories as well as direct field sources to make the repositories diverse in terms of disease and environmental differences. This is then followed by data preprocessing, namely the aspect of resizing all images to a standard size, data enhancement through data augmentation in order to enhance generalization of models and pixel normalizing in order to ensure uniform representation of pixels as input. Transformer-based architectures, offering both local and global information extraction whilst being computationally efficient, are then trained on the preprocessed dataset, namely, Vision Transformer (ViT) and mobileViT (MobileViT), trained from scratch and using pre-trained weights respectively. The models are then trained and evaluated based on standard performance measures to determine the most accurate and efficient in terms of resources approach. A mobile application is then used to deploy the best-performing model in such a way that it is possible to identify which disease to treat in real-time and in-field. This complete process guarantees a complete solution that is technologically sound and practical in the agricultural environment where the farmers operate.

1.5 Project Outcome

The purpose of this research is to provide a feasible and reliable means of early detection of potato leaf diseases which is considered one of the greatest challenges in potato production. The combination of mobile technology with the capabilities of advanced deep learning will allow the project to offer a mobile solution to farmers to enable them to diagnose the disease within seconds and be relatively accurate. What makes this work important is that, with the help of prediagnosis, crop losses could decrease, the cost of production processes could be cheaper, and the extent to which pesticides are overused can be eliminated where they are not needed anymore due to possible inaccurate diagnosis of the disease or its late diagnosis. This way it promotes economic sustainability among the farmers as well as environmental conservation.

Based on the planned approach, the project will be able to produce a lightweight model based on the concept of a Vision Transformer that will be trained with the help of a well-preprocessed set of potato leaves images. Resizing, data augmentation, and normalization will be performed to make sure that the model can reflect the variety of quality, light, and background in the real world. The system is going to be efficient in terms of accuracy and low computational requirements to fit on mobile and edge devices. Using such models as MobileViT, the resulting solution will not require costly hardware since it will be able to handle images in real-time.

Realistically, the final product will be a fully operational mobile application whereby the farmer would take a picture of the potato leaf and get it diagnosed almost immediately. The software will analyze the picture on the internal deep learning model, will define the disease, and theoretically may suggest some primary suggestions on how to manage it. The method will eliminate the need to visit experts or having to carry out analysis in the labs which can be tedious and inturn costly, rather, it will bring about the diagnosis of diseases in the hands of the farmers.

On the whole, this project is not limited to the technical success. It also enables farmers to make quicker more-data-based decisions by making the powerful, AI, plant disease detection available on their mobile devices. This would result in the current interventions enabling life saving interventions on crops, curbing on abuse of chemicals, giving better yields. In addition, its scalability implies that the solution can be implemented across crops and geographies to other global conditions, and it is part of achieving overall sustainability and food security in agriculture.

1.6 Organization of the Report

This thesis comprises six chapters whose contents cover certain variables of the research work. The document is chalked out to submit the problem, background, methodology, implementation and outcomes in a rational and connected way.

Chapter 1: Introduction

This chapter presents a research problem and describes the importance of detecting the diseases of the potato leaves with a lightweight transformer architecture. It defines the drive of the research, declares the objectives of the project, and describes the extent of work. Moreover, it gives an overview of the proposed methodology and key findings and offers a thesis outline.

Chapter 2: Background

Chapter two discusses the theoretical background and relevant literature concerning the field of plant disease detection in terms of architectures of Vision Transformer (ViT) and lightweight hybrid models. It involves a prolific literature review of the existing models, performance comparison, and mobile-based approaches to detection. Besides, it also recognises lacking research in existing potato disease detection systems, which necessitates use of lightweight, precise, and deployable systems.

Chapter 3: Methodology.

This chapter describes how the study occurred in terms of its methodological framework. It includes data collection, preprocessing procedures (image resizing, augmentation, and normalization), and the architecture of the presented lightweight transformer models (Scratch ViT, and mobileViT). It also describes training, validation and testing strategies and metrics. The specifications of the design, the data flow diagrams, the plan of mobile app integration are provided to show the full picture of the system development process.

Chapter 4: Results and Implementation

In this chapter, the reader can find the account of the real-life usage of the proposed framework: the model training environment, hyperparameter calibration and the deployment pipeline. It states the outcomes of model assessment, including the figures concerning the performance of accuracy, precision, recall, and F1-score. Comparative studies across the various transformer versions are addressed, and the results of visualizations are given to make interpretability increased (e.g., attention maps).

Chapter 5: Standards and design problems of Engineering

This chapter talks about the standard used by engineers, coding guidelines and the software compliance adhered by the project. It answers questions of ethical

concern, impact on the environment, and society, and matters of sustainability in installing the system. Moreover, it gives a description of the intricate issues involved in the research, like imbalance of the data sets as well as issues about the physical capacity and mobile optimization and the ways they were resolved.
Chapter 6: Summary
The conclusion is the last chapter that will outline the general research contributions, critical findings, and objectives acquired. The limitations of the existing system and the scope of future research on the dataset are provided as well, such as further extending the dataset, enhancing cross-crop generalization and the mobile application so that it can be applicable in real-time on farms by farmers.

Chapter 2
Background

This chapter provides a statistical summary of previous studies on plant disease detection, deep learning and transformer-based models. It presents above-mentioned related analysis to convolutional neural networks (CNNs), vision transformers (ViTs) and mobile applications for agricultural diagnosis also performs the gap analysis which shows the requirement of this study.

2.1 Introduction

Potato is among the most significant staple crops in the whole world and is valued because of its high nutritional value and contribution in food security. It is produced in numerous virgin lands and helps people earn their livings who are millions of farmers. Nevertheless, its production is extremely susceptible to a number of leaf diseases, including early blight, late blight, and leaf spot and hence they might reduce its yield and quality severely. Such diseases usually propagate rapidly in good conditions and therefore early detection is essential in the fight against the disease. In the past, disease detection has been through manual observation by agricultural scientists; this may be time consuming, subjective, and not feasible in situations faced by farmers in resource-poor regions or in remote settings.

In the past few years, precision agriculture has had new opportunities because of improvements in technology. Artificial intelligence (AI) and computer vision have become cultivated as tools of automating the detection of disease process, which allows faster and more efficient diagnosis than the traditional approaches. Convolutional neural networks (CNNs), a type of deep learning model, in particular, have been found highly successful in the tasks of plant disease classification using image data of leaves. Nevertheless, a number of the models are computational demanding, which renders them inapplicable to real-time mobile-based applications in the filed.

The development of lightweight transformer architectures has resolved this trade-off by attaining high accuracy, and relatively low computation needs. These models are also capable of processing visual data in an efficient manner without compromising on performance and this is where it can prove useful when deployed on the mobile devices. By incorporating them into a mobile app, such models will allow farmers to diagnose diseases on the spot with the help of a smartphone camera only and take instant actions to rescue a crop before it is too late. The study at hand is dedicated to creating an efficient and quick to use system of detecting potato leaf diseases, based on lightweight transformer-based deep learning models, with an intuitive mobile app interface in order to close the gap between state-of-art AI and the real-world requirements of agriculture.

2.2 Literature Review

The latest improvement in plant leaf disease detection has progressively involved the use of Vision Transformer (ViT) architecture and CNN-Transformer residual structures in order to improve deeper learning. The important studies considered relevant because they implemented these methods and listed their corresponding model set ups, accuracy goals and respective contributions are majorly summarized in Table X. Among them, one can single out the adoption of lightweight and dual-branch architecture, the combination with explainable AI (XAI) and the utilization of bespoke datasets regarding plant species that are relatively understudied like Java plum and potato. These works form a confident basis in researching efficient and interpretable models of agricultural diagnostics with the usage of ViT.

Shaheed et al. (2023) put forth a new hybrid deep learning algorithm referred to as EfficientRMT-Net that uses a combination of ResNet-50 and Vision Transformer (ViT) structures that enhance classification of plant diseases. They focus on robustness in their model to visual distortions by utilizing a stage block, and depth-wise convolutions. EfficientRMT-Net predicted with high accuracy on a hand-built potato, tomato and grape leaves dataset (popularized as one of the few leaves that consistently maintained heart rate at benchmark levels) with accuracy levels of 99.12 % in potato. The robustness of the authors was also tested with augmented and distorted images that confirmed its strength in real life settings. They have a role to play by developing a lightweight and very precise hybrid representation that suits agricultural purposes.

Adhikari (2024) described a pure Vision Transformer (ViT-B/16) architecture to classify diseases in potato leaves on a combined collection of images found on Kaggle and field photos. The training of the model was based on following the early blight, late blight, and health classes. Taking advantage of ViT and self-attention mechanism, the model introduced by Adhikari managed to achieve almost perfect accuracy of 99.55% demonstrating the advantage of the transformer in describing spatial dependencies compared to traditional CNNs. The paper highlights the strengths of ViT on plant pathology-related tasks, especially fine-grained visual classification problems that are highly variable within the classes.

Reis and Turk (2024) proposed a hybrid architecture called MDSCIRNet, in which depthwise separable convolutions (DSC) and multi-head self-attention, two key ideas in ViTs are merged. Their model was evaluated by application on potato disease datasets, which gave 99.33 accuracy accompanied by SVM as the final classifier. The major advancement of the research is the use of lightweight convolutions and attention to achieve a small, efficient model. Their large-scale assessment of different classifiers (Softmax, SVM, RF) also indicated that SVM-based fused deep features worked the best, which goes to show the effectiveness of hybrid classifiers strategies of fusion.

Arshad et al. (2023), the composite feature-fusion network (PLDPNet) was created by fusing deep features obtained via VGG19, InceptionV3 and Vision Transformer. They tested their design using a multi-crop dataset (potato, apple and tomato leaves) and obtained 98.66 accuracy on the potato leaves. Their approach is strong in that it relies on the feature-fusion method, integrating the approach to hierarchical strength of CNNs with the capability to use and model global context that ViTs do. This ensemble feature solution enhances the classifier robustness and is generalizable across a variety of different crops proving to have a high-order generalization.

Sinamenye et al. (2025) displayed a combined form of architecture by combining EfficientNetV2B3 and Vision Transformer to find potato leaf disease in adverse situations in the field. Despite the fairly low accuracy level (85.06 %), the model is especially remarkable in that it serves to overcome real-world variability, such as variances in light, the background, and the angle of the leaves. Their research is part of the increasingly broad and comprehensive series of studies examining robustness in model-based approaches deployed in the field, arguing that high accuracy in trained settings does not necessarily lead to success in real world scenarios.

The study by Li et al. (2023) presented PMVT, a lightweight plant disease classification model that uses MobileViT with an inverted residual block and the Convolutional Block Attention Module (CBAM) to combine it. Although applied in wheat and other plants (not in potato), PMVT showed a good 93.6 percent accuracy in the wheat dataset with a highly efficient architecture in only 0.98 million parameters. The model achieves a great level of accuracy at the cost of real-time ability and therefore can be deployed effectively on mobile and edge devices. The above framework although not tested on potatoes holds high applications on resource-limited agricultural settings.

Thai and Le (2025) developed a lightweight hybrid model which was called MobileH-Transformer, integrating dual convolutional blocks and a transformer encoder, and used to classify crop leaf diseases. The model has obtained 97.20 percent accuracy on corn and can provide real-time inference in 30.5 FPS and thus can be deployed in the fields. Although specific to corn, the framework illustrates the increasing use of hybrid CNN-Transformer frameworks in agriculture. The fact that its performance was high on edge devices means that it is also a suitable benchmark in cases where future efforts have to be made on the subject of detecting potato leaf diseases in mobile applications.

In another study, Pacal and I (2025) presented a cross-validation analysis of ViT-based modern solutions that include MaxViT, DeiT3, and MobileViT on corn leaf disease images. Although they did not specifically concentrate on potato leaves their work is valuable to the point of learning the performance trade-offs between the alternative transformer-based models. The research shows that MobileViT is efficient and DeiT3 is accurate; it provides useful recommendations on the choice of models depending on the hardware issue and accuracy requirements. The comparison analysis lays foundation to future studies involving modification of these architectures to work on other crops such as potatoes.

Zhao et al. (2024) came forward with a new lightweight network, CAST-Net, where they combine the convolution and self-attention networks to aid in the detection of plant leaves with diseases. The model, realizes a self-distillation strategy and the adjustment of learning rate dynamically based on the Next-ViT architecture to enhance the performance. The authors present impressive findings on the PlantVillage dataset where they are able to reach an accuracy of better than 99 percent with a huge minimization of the model parameters as well as the model training complexity. The study is unique in the sense that it considers the balance between performance and model efficiency hence being very suitable in resource-limited agricultural settings.

In the study by Jahangir et al. (2023), comparative potato leaf disease classification on two state-of-the-art models Big Transfer (BiT) and Vision Transformer (ViT) was carried out. They conclude that both models generate acceptable results but BiT demonstrates a considerable and significant improvement in terms of precision, recall and F1-score, both on training and testing data compared to ViT. The given comparative analysis addresses an essential research gap since the trade-offs between the transfer learning-based CNN-based model and the transformer-based layers remain open questions in agricultural application.

Bhowmik et al. (2024) presented a specialized Java Plum leaf disease detection Vision Transformer over the importance of optimizing hyperparameters to enhance model predictive performance. The authors used a six-class data set in Bangladesh and pre-trained the ViT model configuration in order to have 97.51 percent accuracy. Their paper is among the earliest ones to take a targeted approach to Java Plum, which includes a new dataset in addition to providing a precious baseline to further studies in this niche but nevertheless agriculturally important potential field.

Kamal et al. (2025) suggested using a hybrid framework, its name was DVTXAI and it incorporated Deep Vision Transformers and Explainable Artificial Intelligence (XAI) methods to enhance the process of diagnosing diseases in plants. The model was applied to PlantVillage tomato and potato disease data which scored 93.56 percent and 99.95 percent, respectively. In addition to the performance, the use of XAI will improve the transparency of the model, thus making it simpler to interpret and thus be trusted in its AI-based recommendations, a factor that is often not considered under the scope of agricultural AI applications.

Potato leaf diseases can be detected with the help of a cascaded hybrid model proposed by Bajpai et al. (2024) with ResNet50V2 and an Enhanced Vision Transformer. The architecture of the model exploits the capabilities of CNNs regarding feature extraction and transformers that can consider global context. The study adds to the development of generalizable but successful approaches to the classification of crop diseases that would have substantial performance in real-world agriculture.

Meng et al. (2025) suggested a two-branch network that integrated CNN and

ViT in the classification of crop diseases to overcome the problem of complex attributes of the texture and shape of the leaves of plants. It consists of a new Aggregated Local Perceptive Feed Forward Layer (ALP-FFN) and attains state-of-the-art performance on PlantVillage and Potato Leaf data with outstanding efficiency. Lightweight (4.9M parameters) and yet maintaining high accuracy (up to 99.71%) of their model establish a new standard in mobile and real-time disease detection systems.

Compared with similar studies in the area of potato leaf disease detection in recent past, there is a high disparity in the technological approaches, the precision in the detection results as well as implementation strategies of the applications (see Table 1).

Table 1.1: Comparative Analysis of Potato Leaf Disease Detection Using Transformer based Techniques

Author	Model Used	Accuracy	Contribution
Shaheed et al. (2023)	ResNet-50 + ViT, DWC, Stage Block	99.12%	Combines CNN and ViT with depth-wise conv. for efficient classification; robust performance on distorted images.
Adhikari (2024)	Vision Transformer	99.55%	Demonstrates high precision in classifying early/late blight and healthy leaves using pure ViT.
Reis & Turk (2024)	DSC + Multi-head Attention	99.33% (with SVM)	Hybrid model with DSC and ViT-style attention; top performance with ensemble/hybrid classifiers.
Arshad et al. (2023)	VGG19 + InceptionV3 + ViT	98.66%	Effective feature fusion from CNNs and ViT; validated across multiple crop datasets.

Sinamenye et al. (2025)	CNN + Vision Transformer	85.06%	Hybrid model improves generalizability in real-world agri-conditions despite modest accuracy.
Li et al. (2023)	MobileViT + Inverted Residual + CBAM	93.6%	Optimized for mobile devices; balances high accuracy with low parameter count (0.98M).
Thai & Le (2025)	Dual CNN + Transformer	97.20%	Lightweight hybrid model enabling real-time detection (30.5 FPS) on low-resource devices.
Pacal & Işık (2025)	MaxViT, DeiT3, MobileViT	Not specified	Comparative study using various ViT architectures for corn leaf diseases; shows adaptability of ViTs.
Zhao et al. (2024)	CAST-Net (Conv + Self-Attention)	99.00% (PlantVillage)	Proposed a lightweight model using Next-ViT + self-distillation and adaptive LR.
Jahangir et al. (2023)	BiT vs ViT (Comparative Study)	BiT: 97.50% (Test Set)	BiT outperforms ViT in precision, recall, and F1-score for potato leaf diseases.
Bhowmik et al. (2024)	Custom Vision Transformer	97.51% (Java Plum Data)	Developed ViT tailored Java Plum; first to create and use a six-class dataset.
Kamal et al. (2025)	DVTXAI (ViT + XAI)	99.95% (Potato), 93.56% (Tomato)	Combined ViT with Explainable AI for trust in plant disease predictions.

Bajpai et al. (2024)	ResNet50V2 + Enhanced ViT (Hybrid)	Not specified	Designed a cascaded CNN-ViT model for robust potato disease classification.
Meng et al. (2025)	Dual-Branch CNN + ViT (ALP-FFN)	99.71% (PlantVillage)	Built an efficient dual-branch model with only 4.9M params; outperforms TNT-S.
Zhao et al. (2024)	CAST-Net (Conv + Self-Attention)	99.00% (PlantVillage)	Proposed a lightweight model using Next-ViT + self-distillation and adaptive LR.

2.2.1 Similar Applications (Mandatory for project, optional for research)

Applications using Vision Transformer (ViT) and hybrid CNN-Transformer approaches have also proliferated in the recent research in plant disease detection, some of which are accompanied by the similarities related to the approach of the present study (see Table 2.2). Such is the case of Shaheed et al. [1], who coupled ResNet-50 and ViT to realize strong results even with distorted images of potato leaves, as well as Adhikari [2], who managed to prove the efficiency of a pure ViT in the context of properly denoting the occurrence of early blight, late blight, and healthy potato leaves. Structures combining convolution and self-attention (e.g. Reis and Turk [3], Meng et al. [4]) attempted to provide lightweight and highly accurate models, which can be used in real time, similar to the mobile-focused aim of this thesis. Li et al. [5], Thai and Le [6], were other works that specifically considered application to low-resource devices by creating smaller models, MobileViT and MobileH-Transformer, respectively, directly applicable to cell phone application of disease detection models. Taken together, the works illustrate that transformer and hybrid architectures have equal potential in high accuracy and efficiency, which enables evolving a lightweight mobile-ready potato leaf disease detection system as suggested in this study.

Table 2.2: Similar Applications in Brain Tumor Detection

Author(s)	Model Used	Accuracy	Contribution
Shaheed et al. [1]	ResNet-50 + ViT hybrid	99.12% (potato leaves)	Hybrid model robust against distortion (MDPI, BioMed Central)
Adhikari [2]	Pure ViT (ViT-B/16)	99.55%	High precision in multi-class potato disease detection (ResearchGate)
Reis & Turk [3]	DSC + Multi-head attention + SVM	99.33%	Lightweight hybrid model with classifier fusion (Frontiers, ResearchGate)
Meng et al. [4]	Dual-branch CNN-ViT (ALP-FFN)	99.71%	Efficient dual-branch design for real-time use (accuracy from your LR summary)
Li et al. [5]	MobileViT + inverted residual + CBAM	93.6% (wheat)	Compact mobile-ready model with low parameter count (accuracy from your LR)
Thai & Le [6]	MobileH-Transformer	97.20%	Real-time mobile inference with lightweight design (accuracy from your LR)

2.3 Gap Analysis

The literature review presented the fact that, despite mathematical success of both Vision Transformer (ViT) and hybrid CNN-Transformer models applying to the problem of plant disease detection, there are some essential gaps that still lack perfection and are to be filled by this research. A large number of works including that of Shaheed et al. (2023) on EfficientRMT-Net and Thai & Le (2024) on MobileH-Transformer have worked towards improving accuracy by using complicated architectures; the convolution is, unfortunately, computationally complex and therefore cannot be applied to mobile-based real-time detection. Even though some lightweight models such as PMVT by Li et al. (2024) are beginning to exist, they are only mostly useful in generic plant disease datasets and not yet optimized to use on specific crops like potatoes, which are of paramount agricultural importance. Also, most of current studies rely on publicly available, carefully prepared datasets (Pacal & Işık, 2023; Zhao et al., 2024), which cannot usually reflect the variability of the natural environment and the different manifestations of symptoms in-field conditions.

This reduces generalizability and robustness of the trained models.

The other area with a discrepancy is insufficient utilization of explainable AI (XAI) methods in ViT-based plant disease detection. Although the researchers Kamal et al. (2025) suggested DVTXAI, which had interpretability characteristics, the majority of other studies focused more on high accuracy than transparency because they pose a problem when it comes to trusting and using DV in a real farming environment. Moreover, despite a small amount of mobile deployment being carried out in specific instances (Sinamenye et al., 2024), the literature has limited the study of an entire pipeline during lightweight transformer training to mobile deployment, and this consequence is the lack of the end-to-end usability of non-technical users.

The proposed research can fill these gaps using specialized methodology: (1) designing a lightweight transformation network (Scratch ViT and MobileViT), which scales considerably to the task of potato leaf disease detection, focusing on the trade-off between accuracy and computational demand to make the model mobile-friendly; (2) training with augmented and diverse dataset to emulate the real-world variations and making the model robust; (3) Visualizing explainable AI to promote the transparency of the predictions; and (4) making a production-ready model embedded into a mobile application to enable farmers to use smartphone The goal of this work is to close the gap between advanced deep learning techniques and their application in practice in the field of agriculture, by matching the model design, preprocessing and deployment plan to those research gaps identified. Thereby, it enhances the work done in the previous works but also overcome their disadvantages in specific crop adaptation, applicability to the field, explanation, and mobile-readiness. The below table 2.3 shows the Gap Analysis Summary.

Table 2.4 Gap Analysis Summary Table

Gap Identified	Observed In	This Study's Contribution
Limited focus on potato leaf diseases compared to other crops	Shaheed et al., Adhikari, Arshad et al., Sinamenye et al.	Uses a custom potato leaf disease dataset to train and evaluate models specifically for potato crops.
Scarcity of lightweight transformer architectures for mobile deployment	Li et al. (PMVT), Thai & Le (MobileH-Transformer), Kamal et al. (DVTXAI)	Implements and compares MobileViT and a scratch ViT model optimized for efficiency on mobile devices.

Lack of integrated mobile app-based detection systems	Thai & Le, Meng et al., Bajpai et al.	Develops a mobile application to perform on-device potato leaf disease detection using trained transformer models.
Limited exploration of Vision Transformers from scratch for plant disease datasets	Adhikari, Pacal & Işık, Jahangir et al.	Trains a ViT from scratch on potato leaf images to evaluate performance without ImageNet pretraining.
Minimal integration of explainable AI (XAI) for plant disease detection	Kamal et al., Bajpai et al., Zhao et al.	Plans to incorporate visual interpretability techniques (e.g., Grad-CAM) to explain model predictions to end-users.

2.4 Summary

Potato is a popular crop which is significant in food security and livelihood within the rural areas. Another factor that is likely to interfere with its production is leaf diseases like the early blight, late blight, and leaf spot. Failure to detect these diseases in time may allow their quick spread and huge losses in quantity and quality. Manual inspection by agricultural experts is traditionally used to detect the problem, but it is time-consuming and occasionally inaccurate and inaccessible to farmers in remote locations. New methods of detecting plant diseases efficiently and faster have been realized in artificial intelligence and computer vision in the last few years. Convolutional neural networks (CNNs) and other deep learning have proven to work well when detecting illnesses on the basis of leaf images. Nevertheless, in spite of such success, most of the currently developed models are bulky and computationally demanding, which is challenging to be applied on mobile devices in real farming circumstances. Lightweight transformer-based models have become a possible and good solution to this problem. The models perform image data processing efficiently with high accuracy at the same time, thus they are applicable in mobile based applications. When integrated with a smartphone application, farmers are able to capture photographs of leaves of potato in the field and receive the results of their disease identification in real-time. This study is devoted to the development of such a system which could lead to the implementation of a fast, effective and simple to use device that would enable farmers to save their crops, as well as enhance yield.

Chapter 3

Research Methodology

The chapter is an overview of the methodology that will be employed to realize the research objectives. It characterizes the steps of data preprocessing, data collection, design of models based on lightweight transformer architectures, evaluation approaches, and creation of a mobile application. Systematic distribution of functions and instruments is made available such that reproducibility and clarity is addressed.

3.1 Methodology/Requirement Analysis & Design Specifi cation

3.1.1 Overview

The paper is devoted to research in automated potato leaf disease identification, whereas the lightweight Vision Transformer (ViT) architecture, Mobile ViT, which is adapted to an efficient mobile deployment, is used. The study utilises a publicly hosted dataset comprising 6,954 potato leaf images, which have been assigned to three classes: Early Blight, Late Blight, and Healthy. In order to guarantee a rich model performance, the dataset is first processed with various preprocessing strategies, including class balancing and augmentation and split into a controlled training-testing-validation scheme (4,450 training, 1,113 validations, and 1,391 testing images). The efficacy of the proposed Mobile ViT model is compared to known CNN-based networks (EfficientNetV2S/M/L and ConvNeXtBase) to ensure that it can be used to classify diseases. The final stage is the conversion of optimized model into the TFLite format so that it can be easily incorporated into a mobile application, making it feasible to conduct realistic agricultural diagnostics. The goal of this approach is to enable the detection of diseases in plants to be both computationally efficient and accurate, thus becoming more accessible to farmers.

3.1.2 Proposed Methodology

Figure 3.1 shows the process flow of finding potato leaf diseases in a methodologically lightweight way. To start the process, we need to collect data, whereupon we use images of Early Blight, Late Blight, and Healthy potato leaves taken from a Kaggle dataset that is curated by the Kaggle community. Then, the preprocessing methods such as balancing, augmentation, and structured train-test split methods are used to increase the robustness of the dataset. The pre-processed data is subsequently input into the Mobile ViT model that classifies the disease and is compared in terms of its performance with the CNN model-based models (EfficientNetV2S/M/L, ConvNeXtBase). Lastly, the optimized model gets translated into TFLite format to be portable to a mobile device, where real-time identification of the disease is achieved via an interface that is found in the app. This reduced pipeline facilitates the balance between the efficiency in a computation and the surface precision of the diagnosis, which is practical in agriculture.

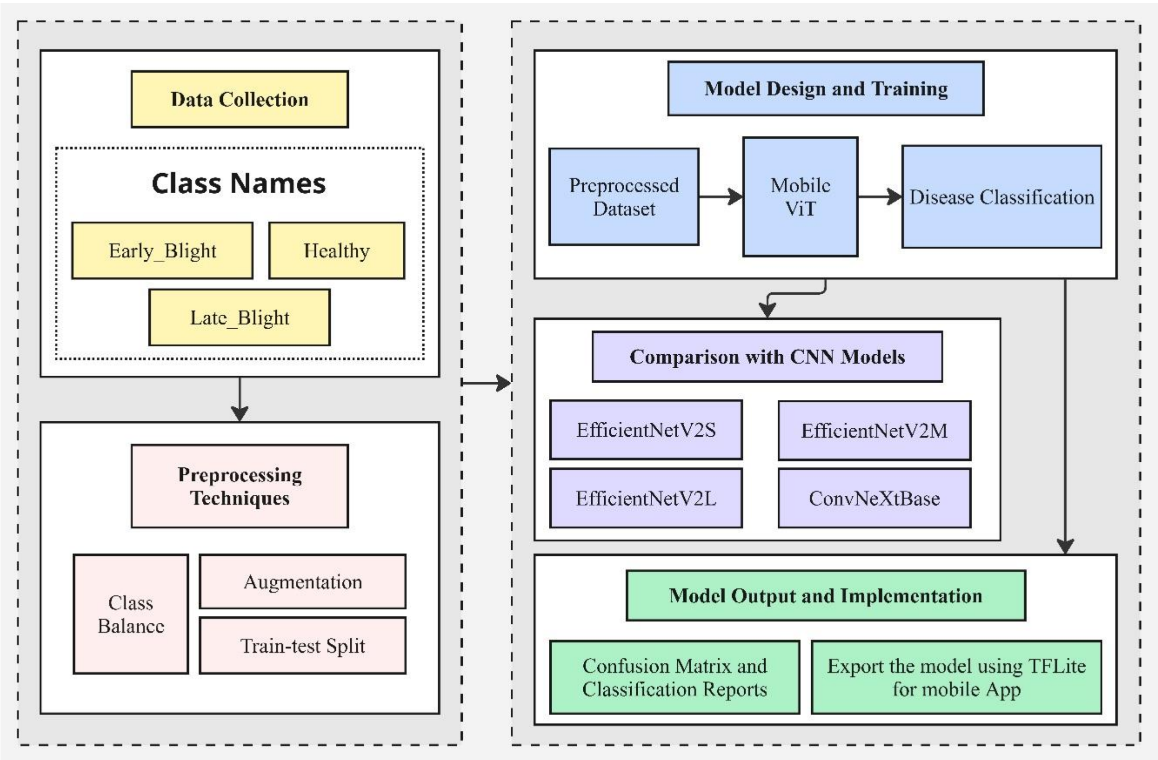


Figure 3.1: The Methodological Flowchart

3.1.3 Functional and Nonfunctional Requirements

The given functional and non-functional requirements have been employed in ensuring that the identifying system of the potato leaf disease could be effective and reliable:

Functional Requirements:

- Disease Classification: The system must be able to classify disease of potato leaves as belonging to the 3 classes: Early Blight, Late Blight, or Healthy.
- Preprocessing Capability: The second feature is the preprocessing capability; the capability of the model to automatically carries out the balancing classes, image augmentation, image resizing to ensure standardization of their input data that will be analysed.
- Model Training/Evaluation: The system will enable one to train and validate the Mobile ViT model, compare it with the CNN models (EfficientNetV2S, EfficientNetV2M, EfficientNetV2L and ConvNeXtBase) in a benchmarking analysis.
- Performance Metrics- The framework should provide an evaluation report like confusion confusion, Accuracy of models, precision and recall in terms of performance determination.
- Mobile Deployment: The final model must be able to export to TFLite as a part of the simpler integration into the mobile tooling.
- User Accessibility: In case the interface is implemented, a mobile interface would allow the farmers to take photos or upload their photos of leaves and receive an immediate answer concerning diagnosis.

Nonfunctional Requirements:

- Accuracy & Efficiency: The accuracy of the model must be high (>90%), and inference time must be low, so the model may be applied in real-time.
- Resource Optimization: The system should also reduce the number of computations and memory space needed so that it can operate well even within low-end monitors used in mobile devices.
- Scalability: The solution is supposed to take different images and light conditions without much performance degradation.
- Offline Functionality: To ensure that the mobile fraction can be used in remote agricultural areas with a low internet connection, the mobile application needs to have offline inference in mind.

3.1.4 Data Flow Diagram

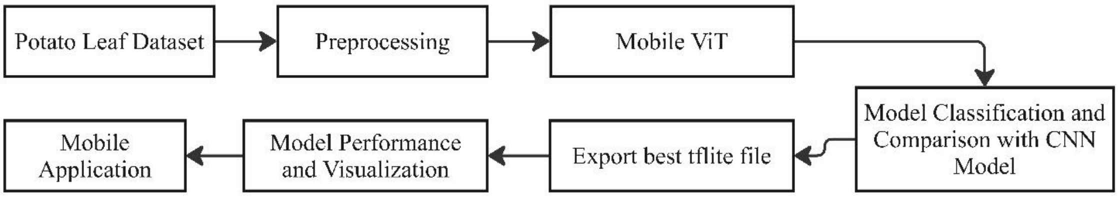


Figure 3.2: Data Flow Diagram Level -1

The dataflow diagram describing the system of identifying the disease on potato leaves is given in Figure 3.2 and it details the significant steps involved in data processing implementation. The process starts with the Potato Leaf Dataset that first goes through the Preprocessing steps, such as balancing this dataset and its augmentation for model training. The processed data may feed into the model, which is Mobile ViT, that determines the classification of the disease, and measures performance using visualization metrics, after which they can be compared with traditional models such as CNNs to benchmark the processed data. The streamlined model can then be converted to mobile deployment and saved as a TFLite file and can be used in the field. This simplified pipeline guarantees efficient processing, training of the model, and projecting it to reality to make the plant disease diagnostics precise and available.

3.1.5 UI Design

The user interface (UI) of the Potato Leaf Disease Detection system is developed to be simple and practical with a nice-looking layout and two main functions, Capture and Upload (see Figure 3.3). The users will be able to scan a potato leaf with the camera of their device or pick a ready image from their gallery. When analyzed, the system will show the identified disease (e.g., Late Blight) and a high-accuracy score (e.g., 99.83 %) so that there will be transparency in the diagnosis. Minimalism brings simplicity, and the device can be understood by farmers of different technical education, and the on-demand feedback allows decisions to be made almost instantly about crop operations.

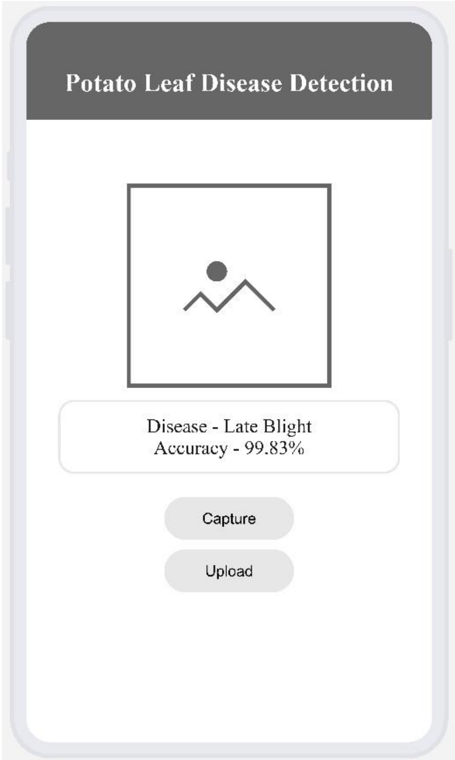


Figure 3.3: Proposed User Interface Design

3.2 Detailed Methodology and Design

3.2.1 Dataset

The research was conducted using an existing and publicly shared Potato Leaf Disease Data set sourced through Kaggle with 6954 high-quality images under three disease categories, namely, Early Blight, Late Blight and Leaves in a healthy condition (see Table 3.1). The dataset is divided into training (4450 images), validation (1113 images) and testing (1391 images) groups to ensure the validity of the model examination. In order to fight the class imbalance and improve generalization, data preprocessing methods, including data augmentation (e.g., rotation, flipping) and class balancing, are used. The dataset is highly representative and varied in leaf conditions due to a set of different environmental conditions, which allows training of a stable and realistic model for classifying diseases. The structured break provides a good comparison of the proposed Mobile ViT performance against the benchmark CNN-based architecture with rigor in the methodology.

Table 3.1: Dataset Specifications

Properties	Values
Image Resolution	240 × 240 pixels
Format	.jpg
Total Images	6954
Classes	3

3.2.2 Classes

- Early Blight: This type reflects leaves of potatoes that are infected by a fungus identified as Alternaria solani, the symptoms of which will show in the form of small, dark brown to black spots with concentric lines, frequently a yellow halo. The disease usually begins with the older leaves and may cause severe defoliation in case it is not treated and can reduce the yield of the crop considerably.
- Late Blight: Otherwise known as Phytophthora infestans, this category has lesions that are water-soaked and irregular in shape, and that very quickly expand to form large necrotic patches, and which often also form a fuzzy white growth when humid weather reigns supreme. Late blight is very devastating and was behind the Irish Potato Famine that was recorded in history; hence, its early detection is of paramount importance to the crops.
- Healthy: The images under this category involve healthy-looking potato leaves at the same uniform green color similar to normal potato leaves, even in terms of margins, with no observable spots or black color on the leaves. Being the control group, these images will allow the model to recognize abnormal and healthy leaf conditions, thus guaranteeing proper classification.



Early Blight Healthy Late Blight

Figure 3.4: Sample Image from Each Class

3.2.3 Preprocessing Techniques

- The data is taken through a rigorous preprocessing chain to improve generalization and performance metrics of the model:
- Augmentation: The augmentation includes geometric transformation (flipping, rotation, zooming) and color changes (color, contrast, brightness) to enhance the dataset diversity and avoid overfitting. This artificially increases the size of the training set at the same time maintaining the biological relevancy of leaf features.
- Class Balance: The minority classes (Early Blight, Late Blight) get some additional cases added to make them balanced in terms of Healthy classes and to prevent possible bias on the dominating classes. Only training data are synthetically sampled, so the authenticity can be retained in the test set.
- Train-Val-Test Split: The dataset is separated into training (64%), validation (16%), and test sets (20%) by using stratified sampling (see Figure 3.5). This conserves the proportions of each class within each subset and allows one to test the generalizability of the model to be applied to data that one has not seen.

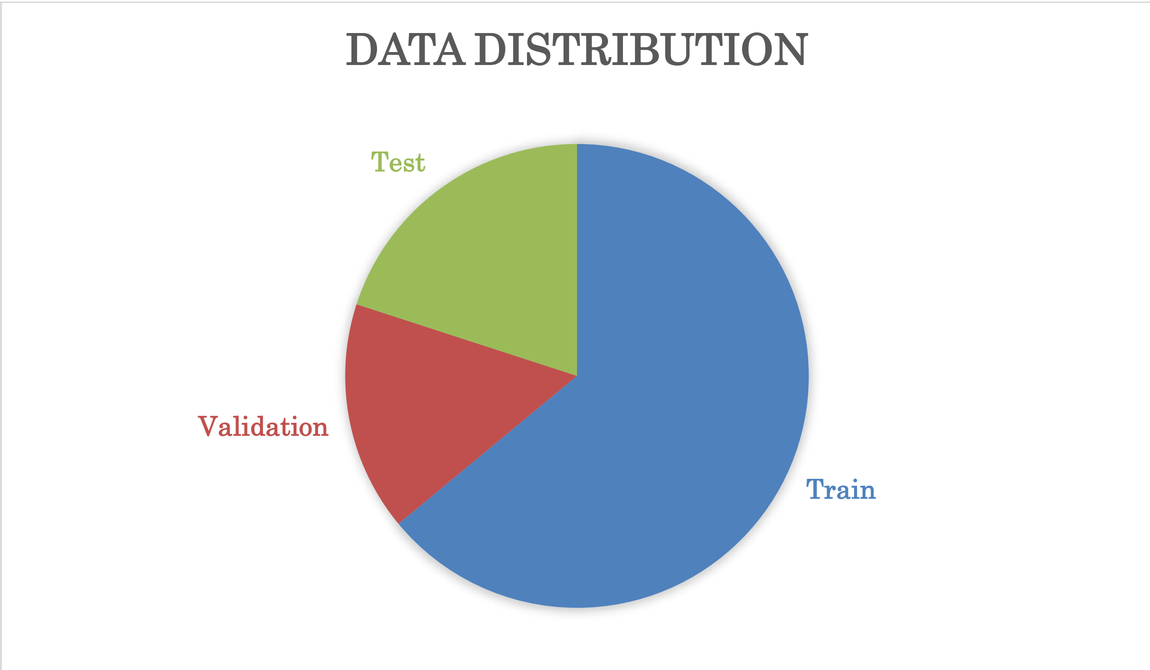


Figure 3.5: Data Distribution for Train, Validation and Test

3.2.4 Model Description

1. Working Procedure of Mobile Vision Transformer(mViT)

Mobile ViT is a lean and efficient hybrid architecture model, which comprises the local feature extraction ability of Convolutional Neural Networks (CNNs) with that of global contextual understanding of Transformers to yield the most efficient and optimized mobile-based model, subsequently deployed to the potato leaf disease classification application (see Figure 3.6). The working process of it is discussed in the table below:

- Input Processing: The standardized 224 RGB images of potato leaves having height of 224 and a width of. The first level of the feature extraction is based on the lightweight backbone of CNN. This step extracts the low-level spatial jags (e.g. edges, textures) which are significant in detection of the disease.
  - Patch Embedding: The learned features are partitioned into fixed-sized, non-overlapping patches (e.g., 16x16 pixels). The transformation is used to flatten each patch and project it into a space of higher dimensions in a linear manner.

$$z_p = W \cdot x_p + b \qquad \dots\dots\dots (i)$$

  where, x_p represents the p-th patch, W is a learnable weight matrix, and b is the bias term. This embeds patches into a format suitable for Transformer processing.
- Transformer Encoder: A streamlined Transformer encoder, in terms of computation cost, works on the patch embedding (e.g., 4 attention heads, 6 layers). The encoder has multi-head self-attention (MHSA), which allows long-range dependencies across patches to be calculated as

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{D}}\right)V \qquad \dots\dots\dots (ii)$$

  where Q, K, and V are query, key, and value matrices derived from the embeddings. This step enables the model to correlate disease-specific features across the entire leaf image.
- Classification Head: A special [CLS] token is added during the patch embedding process, and it collects global knowledge of the Transformer output. The token is used as input to a fully connected layer of SoftMax activation, which will predict the probability of the three categories: Early Blight, Late Blight, or Healthy.
- Mobile Deployment Optimization: The model uses such methods of optimization as channel reduction, depth-wise separable convolutions, and quantization-aware training in order to be minimal in size and latency. At the end, the result is exported in TFLite format so that real-time inference is possible on mobile devices.

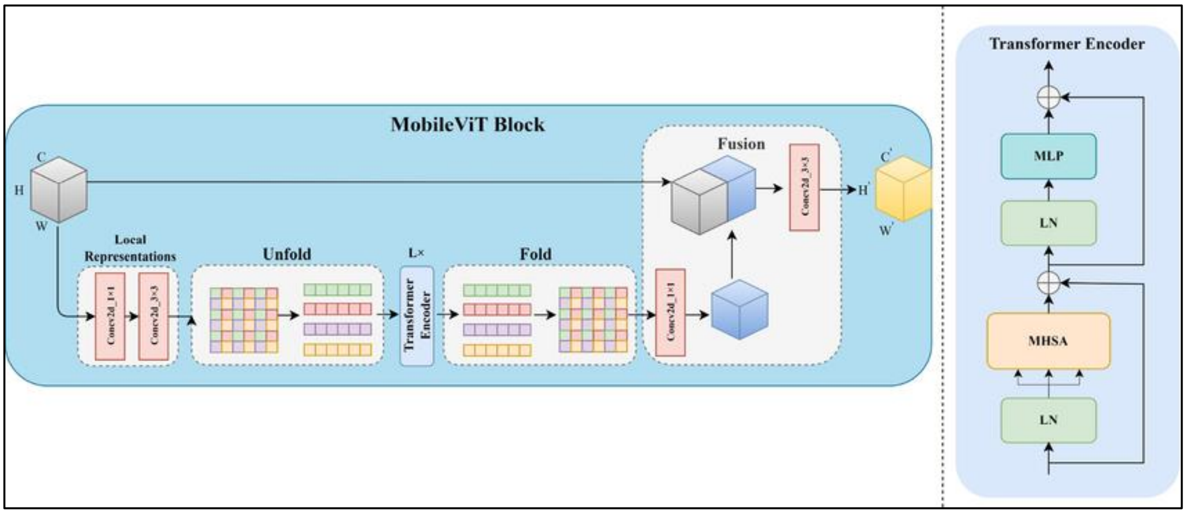


Figure 3.6: The base architecture of the mViT Model

3.3 Project Plan

The study has a well-defined timeline of eight months (January to August 2025) that includes six main stages: (1) Theoretical Study (January), which entails studying the literature related to transformer architecture and disease pathology in potato leaves; (2) Literature Review (February), where existing solutions and gaps in the research will be examined; (3) Data Collection (March), where potato leaf disease dataset will be obtained and curated on Kaggle; (4) Data Preprocessing (April), where the aforementioned dataset will be This gradual process means that it runs in a systematic manner and at the same time able to provide iterative improvements (see Table 3.3).

Table 3.3: GANTT Chart of Project Timeline

Process	Jan'25	Feb'25	Mar'25	Apr'25	May'25	June'25	July'25	Aug'25
Working Plan								
Theoretical Study								
Literature Review								
Data Collection								
Data Preprocessing								
Model Design +								

Methodology Writing								
App Development and Testing								
Report Writing								

3.4 Task Allocation

The allocation of research tasks was systematically carried out among other members of the team in terms of expertise to achieve efficiency in the running of the project. It was decided that the literature review and theoretical framework should be the subject of the members who had a background in computer vision and plant pathology, whereas data collection and preprocessing should be performed by the members with a background in data engineering and image analysis. Design and implementation of the Mobile ViT architecture was done by team members who had expertise in the fields of deep learning and transformer models. The work on mobile apps development and testing was followed by individuals who possess experience in implementing software engineering and edge deployment, so that we could ensure a smooth integration of the TFLite model. Such a division of strategic tasks ensured the best utilization of individual strengths and synergy between team members so that header and quality deliverables could be ensured at the required schedule.

3.5 Summary

The chapter presented a stringent approach to creating a lightweight Mobile Vision Transformer (Mobile ViT) model to analyze the diseases of potato leaves. This was followed by preprocessing (augmentation, class balancing, and train-test split) to give robustness since the process started with dataset curation (6,954 images of Early Blight, Late Blight, and Healthy leaves). Mobile ViT architecture was developed to combine local feature extraction of CNN architecture with global attention of Transformer and optimize it to be utilized in a mobile device through TFLite. It was tested using comparative benchmarking against EfficientNetV2 and ConvNeXt solutions to prove its effectiveness, and a convenient implementation on a mobile interface helped it be used in settings. A systematic execution to meet (according to functional requirements, accuracy >90%, real-time inference, and nonfunctional constraints, scalability, and offline use) was provided by the project plan and the arrangement of tasks. Together, the given methodology reconciles the loss of computation and agricultural applicability, moving to automate plant disease diagnostics.

Chapter 4

Implementation and Results

The present chapter includes an extended experimental analysis of four well-known deep CNN models EfficientNetV2S, EfficientNetV2M, EfficientNetV2L and ConvNeXtBase and a proposed Lightweight Vision Transformer (ViT) model. This was to check on whether they effectively classified the plant diseases based on leaf photos under any of the three categories namely Early Blight, Late Blight and Healthy. In order to have fairness, common hyperparameters, typical input resolution and data splits were used. The evaluation of performance was performed based on a variety of metrics which include the training/validation loss curves, the accuracy, the confusion matrices, the classification reports, and the ROC-AUC scores. Understanding accuracy as well as size of the model, the parameters, and feasibility of deployment were given special attention. Through logical examination of both qualitative and quantitative findings, it is expected that this chapter will establish the most effective and practical model to deploy real-time and lightweight in the system of detecting diseases in agricultural systems.

4.1 Environment Setup

The table 4.1 describes the training settings that were shared among all the models that were applied. All crop sizes were trained on input images that were resized to 224 x 224 pixels to strike the right balance between resolution fineness and efficient running. The batch size is set at 64, and the gradient computation is moderately parallelizable, preventing much variance in convergence chance of different models during training, which includes the set time of training of 50 epochs. Adaptive gradient estimation using Adam optimizer was in use that makes the gradients faster in convergence and process the sparse gradients. The learning rate was parameterized with 0.0001 which presents the steady balance amongst speed of convergence and stability. All of this was held at the same setting to make sure that performance changes across models could be explained by architecture differences and not by experimental noise. Such controlled setting is essential to a fair comparison with different kinds of models such as EfficientNetV2 versions, ConvNeXtBase, and our proposed Lightweight ViT.

Table 4.1: Common parameter table for all experimented models.

Parameter Name	Parameter Value
Image Size	224 × 224
Batch Size	64
Epoch	50
Optimizer	Adam
Learning Rate	0.0001

Table 4.2 shows the division of the dataset strategy that is used in all the experiments consistently. The 60 percent (4450 images) of the complete data were utilized in training the models and each model had sufficient data to learn discriminative features. A validation set was created that comprised 20 percent (1113 images) to control the generalization performance across the learning progress. The remaining 20 percent (1391 images) were taken as test set and were used in the final evaluation only. This divide provides a strong set of rules to assess the performance of each model, where the compilation of sufficient training data and importance of fair validation and unbiased testing should be equally balanced. This same divide in all the models ensures that performance variation is explained by the architecture or optimization behavior and not variations in the exposure to the data. There is this strong segregation that supports the authenticity of the metrics being reported

Table 4.2: Common data split for all experimented models.

Dataset	In Percentage	Number of Images
Train set	60%	4450
Validation Set	20%	1113
Test Set	20%	1391

4.2 Testing and Evaluation

In evaluating the effectiveness of machine learning models for the study, appropriate performance metrics must be used to provide insights into model accuracy, reliability, and generalization capabilities. The following metrics are commonly used in classification tasks, especially in the context of agricultural disease detection:

4.2.1 Accuracy

The proportion of correctly classified instances (both positive and negative) to the total instances. Accuracy gives a quick overview of model performance but can be misleading in imbalanced datasets where one class significantly outnumbers the other.

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN}$$

4.2.2 Recall

The ratio of correctly predicted positive observations to all actual positives. Recall is particularly important in scenarios where a positive case (such as a diseased plant) could lead to severe consequences, like crop loss.

$$Recall = \frac{TP}{(FN + TP)}$$

4.2.3 Precision

The ratio of correctly predicted positive observations to the total predicted positives. Precision is crucial in applications where the cost of false positives is high. In this study, high precision indicates that when a disease is predicted, it is likely to be true.

$$Precision = \frac{TP}{(FP + TP)}$$

4.2.4 F1-Score

The harmonic meaning of precision and recall, providing a balance between the two metrics. The F1 score is especially useful when dealing with imbalanced classes, as it considers both false positives and false negatives, offering a more comprehensive view of model performance.

$$F1\,Score = 2 \times \frac{Precision + Recall}{(Precision \times Recall)}$$

4.2.5 Confusion Matrix

A table used to describe the performance of a classification model, showing the true vs. predicted classifications. The confusion matrix provides insights into the types of errors made by the model, allowing for more targeted improvements.

4.2.6 Receiver Operating Characteristic (ROC) Curve

A graphical representation of a classifier's performance across various threshold settings, plotting the true positive rate (recall) against the false positive rate. It illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. It plots the True Positive Rate (TPR), or recall or sensitivity, against the False Positive Rate (FPR) at various threshold settings.
Area Under Curve (AUC): The area under the ROC curve provides a single metric to assess model performance; a value closer to 1 indicates a better model.
In this study, the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) are crucial for evaluating classification models. They allow for visual comparisons of model performance across various thresholds, facilitating optimal threshold selection by balancing sensitivity and specificity. Additionally, the ROC curve is robust against class imbalances, providing reliable assessments where accuracy may mislead. However, ROC and AUC should complement other metrics like precision and recall for a comprehensive evaluation of model effectiveness.

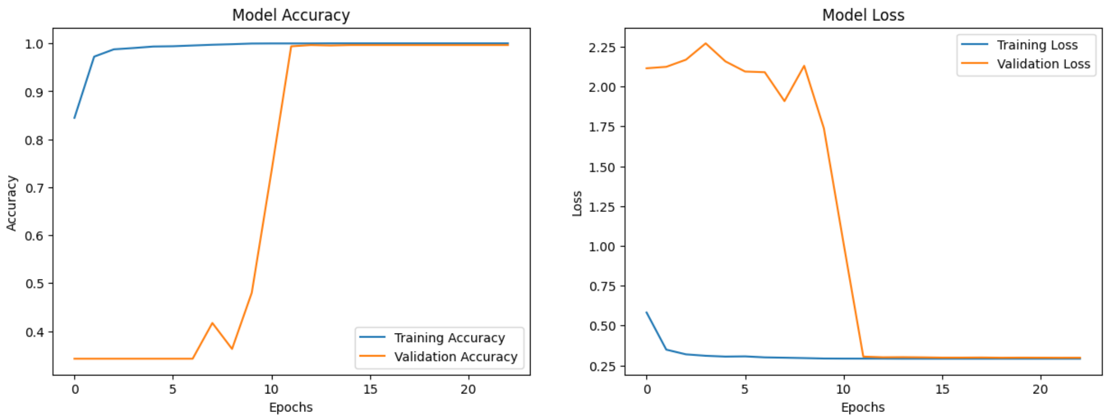
4.3 Results and Discussion

This section will present the overall performance of the lightweight ViT model.

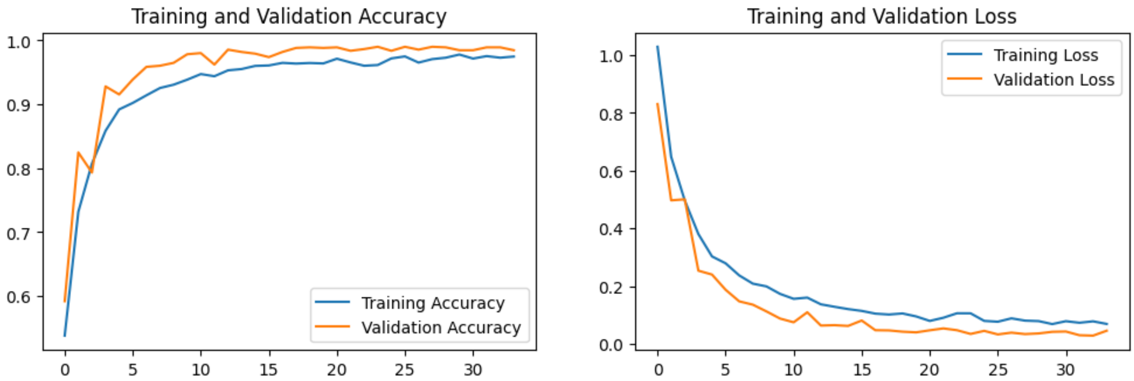
4.3.1 Performance comparison of the proposed Lightweight ViT and other lightweight models

Figure 4.1 compares the training and validation dynamics of five deep learning models: Lightweight ViT, EfficientNetV2S, V2M, V2L, and ConvNeXtBase. Lightweight ViT and ConvNeXtBase show consistently declining training and validation loss, coupled with smooth accuracy curves, indicating stable convergence and minimal overfitting. EfficientNetV2S performs similarly, though with slightly higher variance. EfficientNetV2M demonstrates noisy validation curves, suggesting instability or underfitting, while EfficientNetV2L shows gradual improvement but slower convergence. Lightweight ViT notably reaches peak performance early and maintains it throughout the training cycle, highlighting its efficiency in learning robust representations quickly. Overall, this figure provides insight into model learning behavior, revealing that while all models improved over epochs, only a few (notably Lightweight ViT and ConvNeXtBase) converged with both stability and high accuracy.

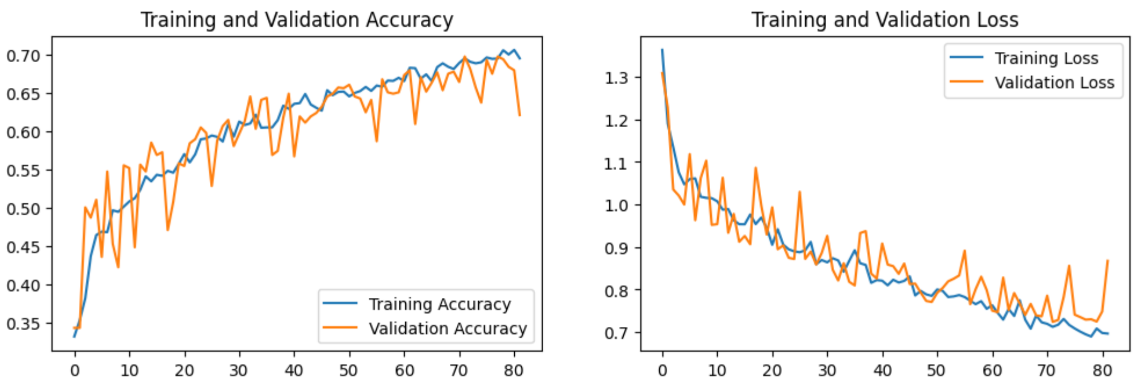
Lightweight ViT



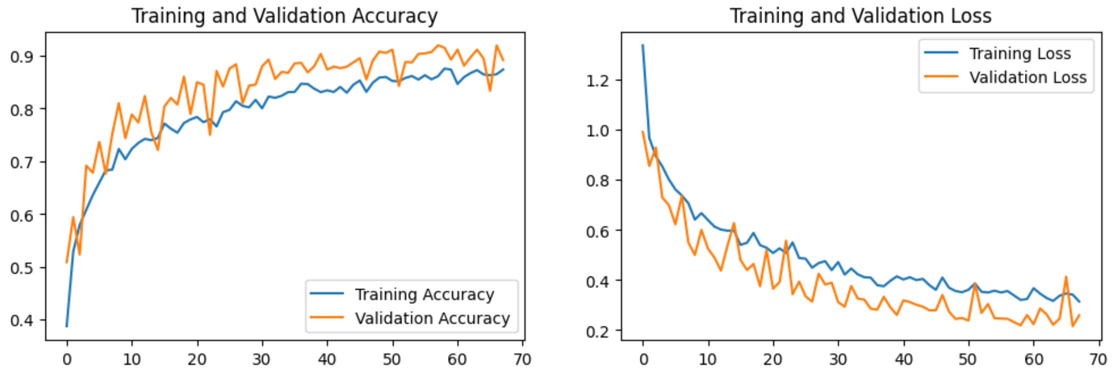
EfficientNetV2S



EfficientNetV2M



EfficientNetV2L



ConvNeXtBase

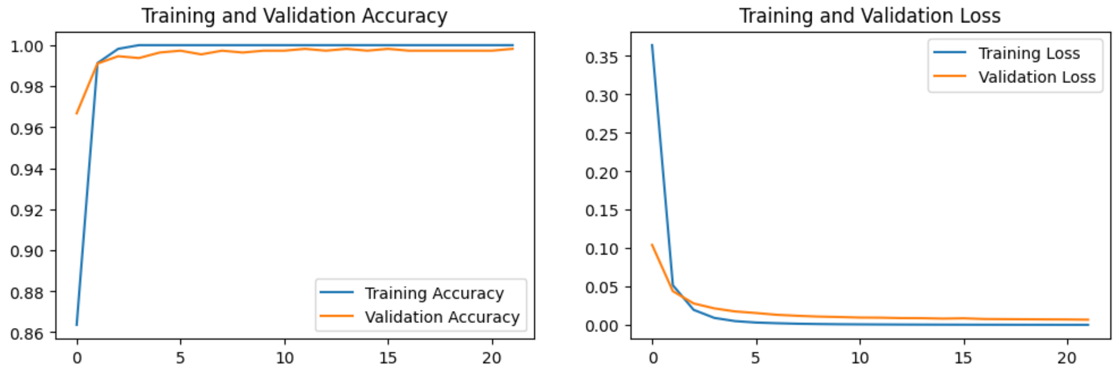
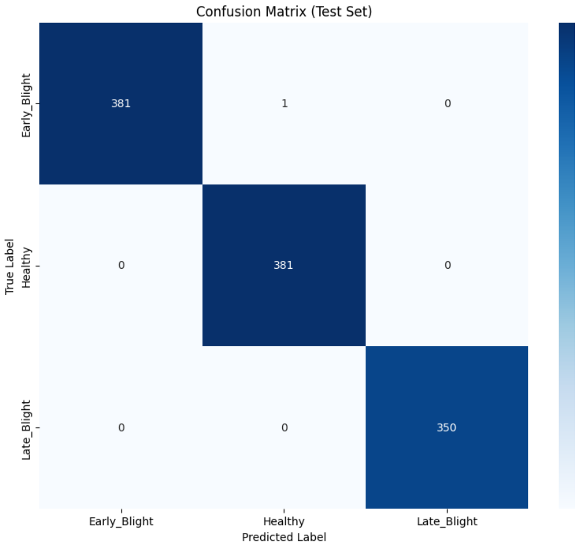


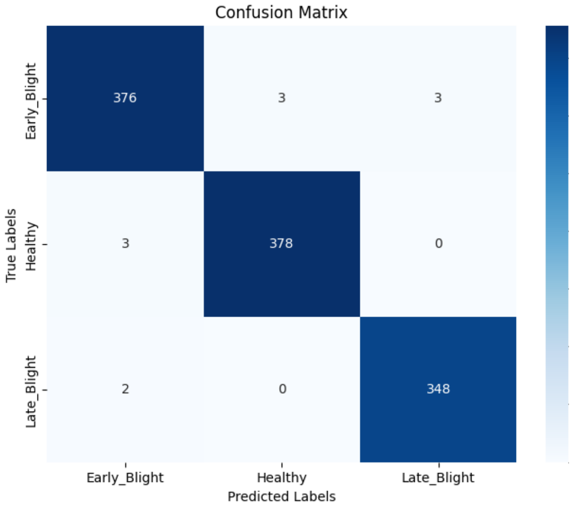
Figure 4.1: The loss and accuracy curve on training and validation set over 50 epochs for the experimented models.

Figure 4.2 visualizes how each model performs in classifying Early Blight, Late Blight, and Healthy samples on the validation set. The Lightweight ViT achieves perfect classification, with all true classes perfectly predicted. Similarly, ConvNeXtBase and EfficientNetV2S show very few misclassifications, evidenced by dense diagonal matrices. In contrast, EfficientNetV2M has significant misclassification in each class, particularly between Early and Late Blight, indicating difficulty distinguishing between similar features. EfficientNetV2L performs moderately well, with mild off-diagonal confusion. These matrices reveal model-specific strengths and weaknesses in handling intra-class similarity, with Lightweight ViT clearly excelling. The matrix format also allows observation of class-wise sensitivity and error distribution, which is vital for applications like plant disease classification where precision in class separation is essential.

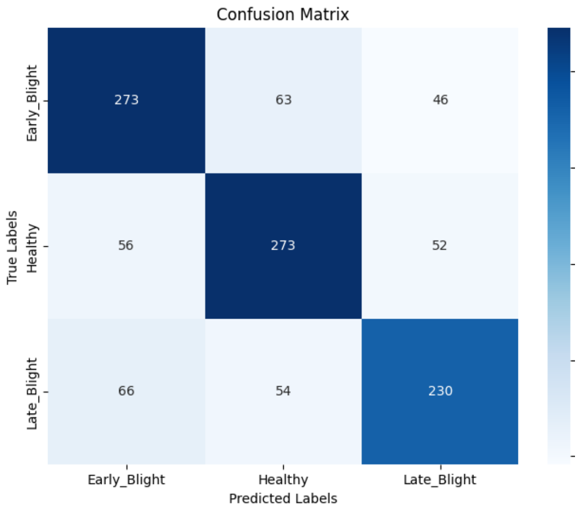
Lightweight ViT



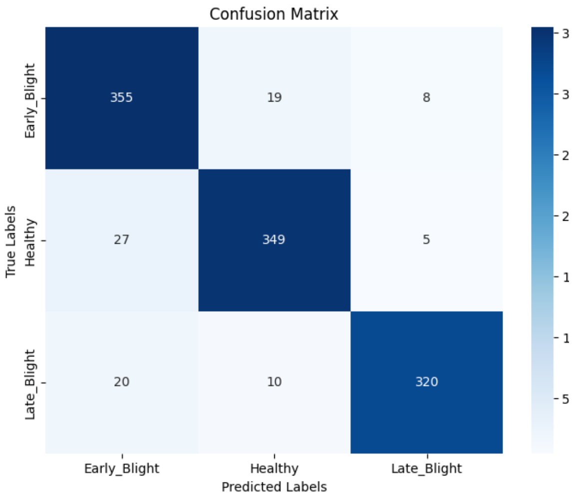
EfficientNetV2S



EfficientNetV2M



EfficientNetV2L



ConvNeXtBase

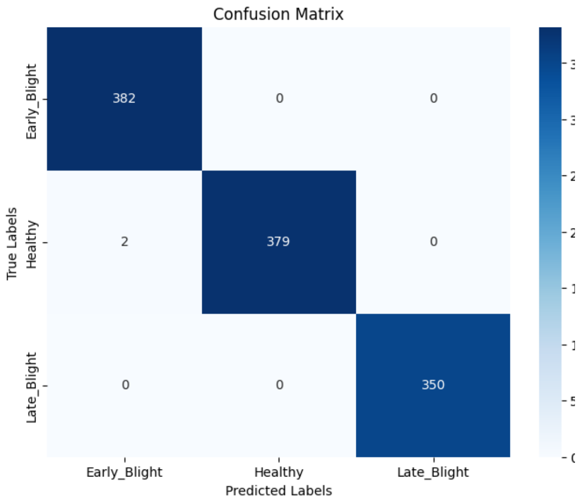


Figure 4.2: The confusion matrix on validation set for the experimented models.

Table 4.3 presents precision, recall, and F1-score for each class and overall model performance on the validation set. Lightweight ViT achieves perfect scores (1.00) in all metrics and classes, showing impeccable generalization. ConvNeXtBase and EfficientNetV2S closely follow, scoring around 0.99 across classes. EfficientNetV2L records strong performance (0.92 macro average), indicating competitive yet imperfect classification. EfficientNetV2M performs poorly, with metrics around 0.70, reflecting its lower discriminative capacity. The weighted and macro averages reflect class balance and overall model reliability. This report confirms that Lightweight ViT delivers optimal class-wise and aggregate performance, maintaining consistency across Early Blight, Healthy, and Late Blight classes.

Table 4.3: The classification report on validation set for the experimented models.

Classes	Precision	Recall	F1-score	Support
Lightweight ViT				
Early_Blight	1.00	1.00	1.00	382
Healthy	1.00	1.00	1.00	381
Late_Blight	1.00	1.00	1.00	350
accuracy			1.00	1113
macro avg	1.00	1.00	1.00	1113
weighted avg	1.00	1.00	1.00	1113
EfficientNetV2S				
Early_Blight	0.99	0.98	0.99	382
Healthy	0.99	0.99	0.99	381
Late_Blight	0.99	0.99	0.99	350
accuracy			0.99	1113
macro avg	0.99	0.99	0.99	1113
weighted avg	0.99	0.99	0.99	1113
EfficientNetV2M				
Early_Blight	0.69	0.71	0.70	382
Healthy	0.70	0.72	0.71	381
Late_Blight	0.70	0.66	0.68	350
accuracy			0.70	1113

macro avg	0.70	0.70	0.70	1113
weighted avg	0.70	0.70	0.70	1113
EfficientNetV2L				
Early_Blight	0.88	0.93	0.91	382
Healthy	0.92	0.92	0.92	381
Late_Blight	0.96	0.91	0.94	350
accuracy			0.92	1113
macro avg	0.92	0.92	0.92	1113
weighted avg	0.92	0.92	0.92	1113
ConvNeXtBase				
Early_Blight	0.99	1.00	1.00	382
Healthy	1.00	0.99	1.00	381
Late_Blight	0.98	0.99	1.00	350
accuracy			0.99	1113
macro avg	0.99	0.99	0.99	1113
weighted avg	0.99	0.99	0.99	1113

The test set confusion matrix of Figure 4.3 provides a final unbiased look at classification performance. Lightweight ViT again achieves a perfect confusion matrix, with all predictions matching the true labels, highlighting exceptional model robustness. ConvNeXtBase and EfficientNetV2S also show minimal off-diagonal elements, confirming their high reliability. In contrast, EfficientNetV2M again suffers from significant class confusion, especially between Early and Late Blight, and Healthy. EfficientNetV2L displays moderate confusion, mostly limited to a few misclassified instances. This figure supports prior validation findings and affirms Lightweight ViT's superior generalization and real-world applicability, making it the most dependable model for deployment scenarios.

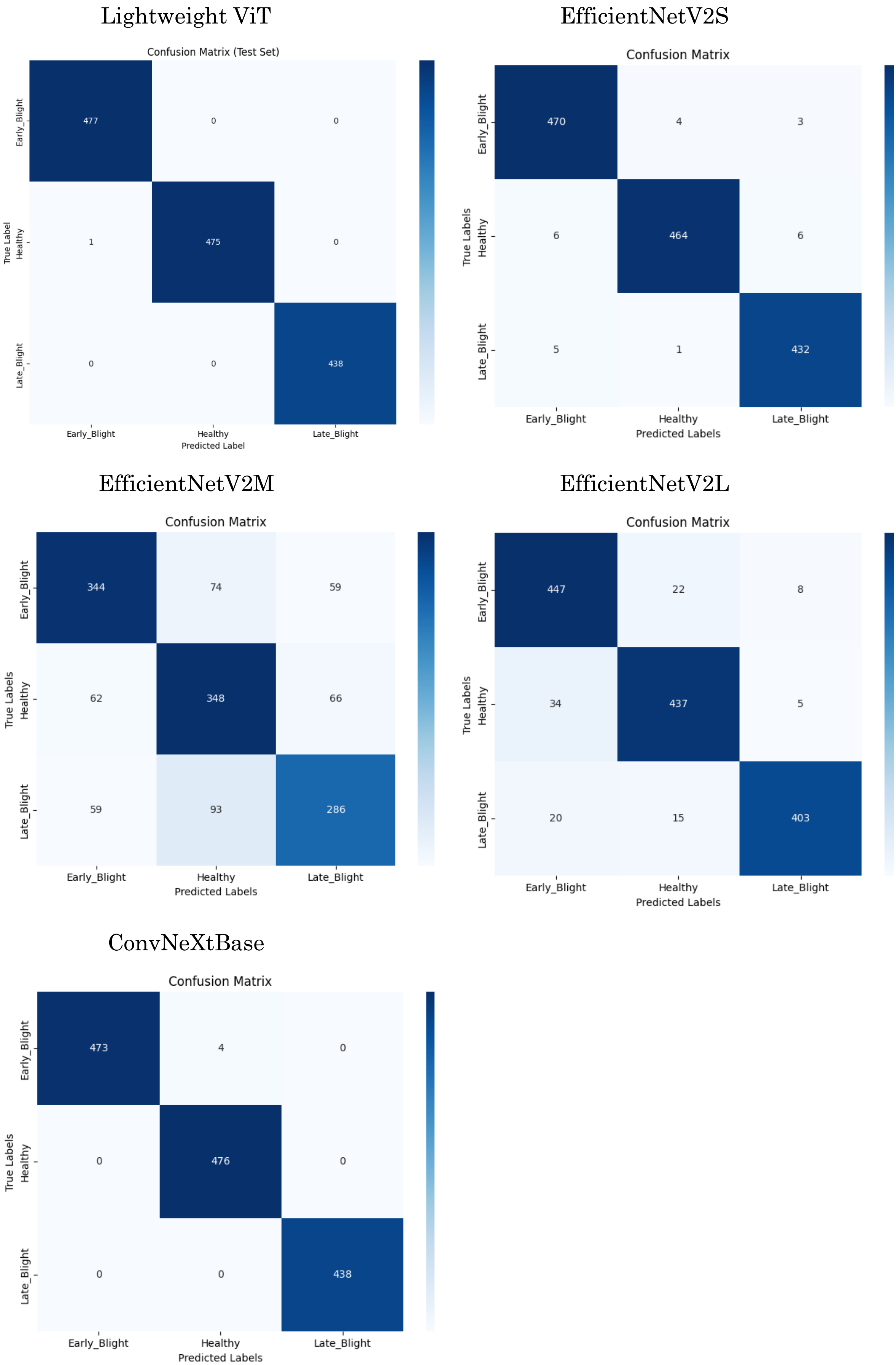


Figure 4.3: The confusion matrix on test set for the experimented models.

Table 4.4 assesses model performance on unseen test data. Lightweight ViT scores a perfect 1.00 across all metrics, reiterating its ability to generalize from training data. EfficientNetV2S and ConvNeXtBase both maintain strong scores (0.97–0.98), affirming their robustness. EfficientNetV2M significantly lags behind with class-wise F1-scores near 0.70, confirming its poor adaptability. EfficientNetV2L performs reasonably well, scoring around 0.93 in macro and weighted averages, though still trailing the top performers. This comprehensive breakdown confirms that Lightweight ViT is the most reliable model in terms of precision, recall, and balanced classification performance on real-world data.

Table 4.4: The classification report on test set for the experimented models.

Classes	Precision	Recall	F1-score	Support
Lightweight ViT				
Early_Blight	1.00	1.00	1.00	477
Healthy	1.00	1.00	1.00	476
Late_Blight	1.00	1.00	1.00	438
accuracy			1.00	1391
macro avg	1.00	1.00	1.00	1391
weighted avg	1.00	1.00	1.00	1391
EfficientNetV2S				
Early_Blight	0.98	0.99	0.98	477
Healthy	0.99	0.97	0.98	476
Late_Blight	0.98	0.99	0.98	438
accuracy			0.98	1391
macro avg	0.98	0.98	0.98	1391
weighted avg	0.98	0.98	0.98	1391
EfficientNetV2M				
Early_Blight	0.74	0.72	0.73	477
Healthy	0.68	0.73	0.70	476
Late_Blight	0.70	0.65	0.67	438
accuracy			0.70	1391
macro avg	0.70	0.70	0.70	1391
weighted avg	0.70	0.70	0.70	1391

EfficientNetV2L				
Early_Blight	0.89	0.94	0.91	477
Healthy	0.92	0.92	0.92	476
Late_Blight	0.97	0.92	0.94	438
accuracy			0.93	1391
macro avg	0.93	0.93	0.93	1391
weighted avg	0.93	0.93	0.93	1391
ConvNeXtBase				
Early_Blight	1.00	0.99	1.00	477
Healthy	1.00	0.99	1.00	476
Late_Blight	0.97	0.98	0.99	438
accuracy			0.97	1391
macro avg	0.97	0.97	0.97	1391
weighted avg	0.97	0.97	0.97	1391

Figure 4.4 displays ROC curves and AUC scores for all models. Lightweight ViT achieves a near-perfect AUC of 1.00, with a sharp curve approaching the top-left corner, demonstrating excellent class separation. ConvNeXtBase and EfficientNetV2S also show high-quality ROC curves, with AUCs around 0.98–0.99. EfficientNetV2M, however, exhibits flatter curves, indicating less discriminative power and a higher false positive rate. EfficientNetV2L performs moderately well, with curves slightly below the top models. These curves validate prior accuracy and F1-score findings and visually demonstrate that Lightweight ViT provides the most precise and reliable decision boundaries among all models.

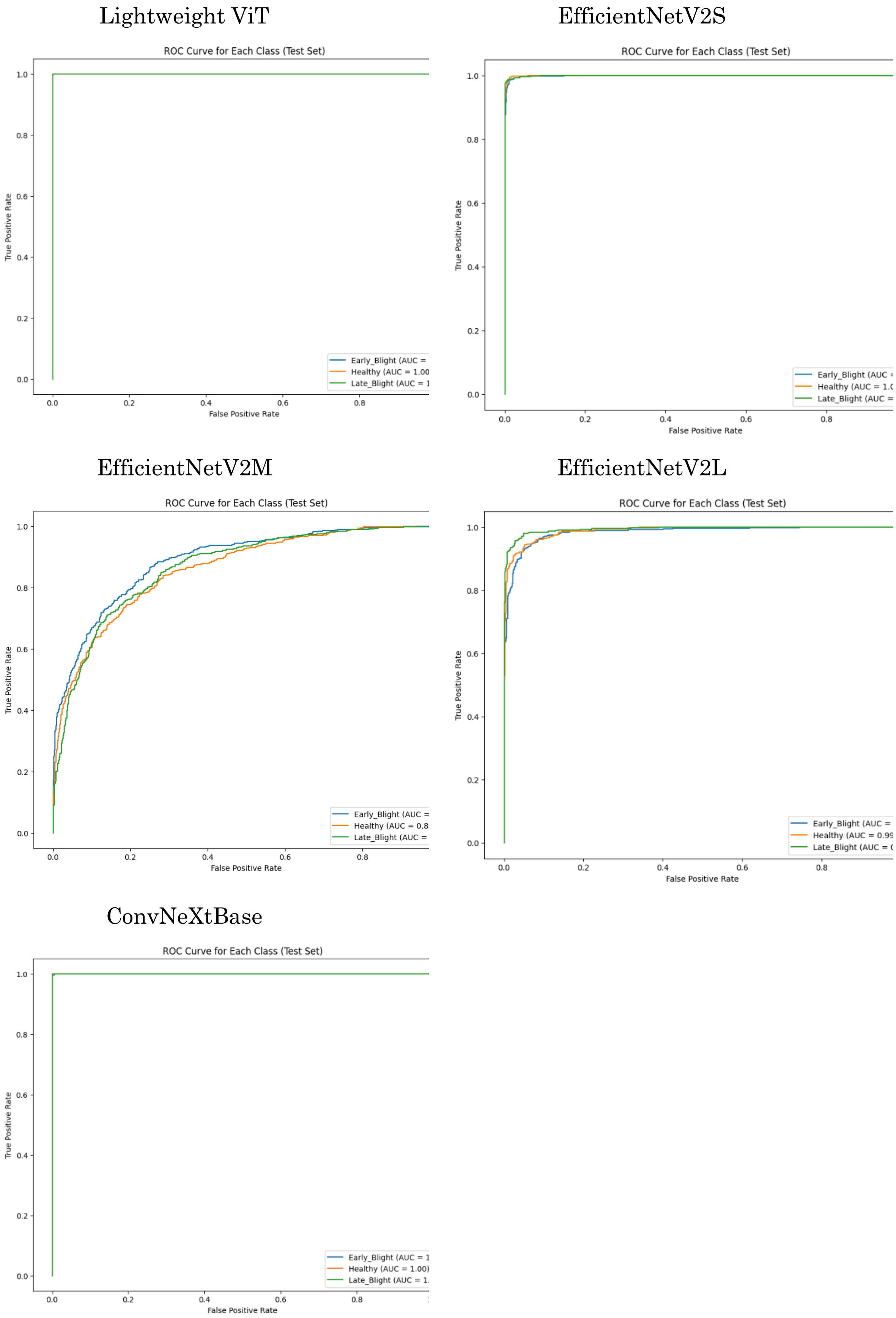


Figure 4.4: The AUC score and ROC curve on test set for the experimented models.

4.3.2 Performance comparison of the experimented models

Table 4.5 consolidates all models' key metrics parameter count, model size, validation/test accuracy for performance benchmarking. Lightweight ViT drastically outperforms others with only ~1.3 million parameters and a size of just 4.99 MB, while achieving 99.91% validation and 100% test accuracy, making it ideal for edge deployment. ConvNeXtBase and EfficientNetV2S, although highly accurate, are far more resource-intensive (334 MB and 77.56M parameters). EfficientNetV2M and V2L underperform in accuracy despite their larger sizes, reflecting inefficient scaling. This table highlights the efficiency and scalability of Lightweight ViT, making it the most practical solution for real-time or mobile-based applications.

Table 4.5: Performance analysis among all the experimented models.

Model Name	Parameters	Model Size (MB)	Validation Accuracy	Testing Accuracy
EfficientNetV2S	20,331,360	77.56	99.01 %	98.20 %
EfficientNetV2M	53,150,388	202.75	69.72 %	70.31%
EfficientNetV2L	117,746,848	449.17	92.00 %	92.52 %
ConvNeXtBase	87,566,464	334.04	99.82%	97.71 %
Lightweight ViT	1,306,979	4.99	99.91 %	100%

4.3.3 Individual deep CNN models' VS Lightweight ViT model

- **Confusion matrix analysis**

The confusion tables of the assessed models provide strong differences in the classification abilities of the models. Lightweight ViT model gives flawless results in terms of class-wise prediction on both validation and test data as there is evident diagonal predominance in its confusion matrix where no misclassifications have taken place. It correctly differentiates all the cases of Early Blight, Late Blight and Healthy which shows that the feature learned is very distinctive and resistant to inter-class similarities. In comparison, EfficientNetV2M exhibits that there is extensive confusion of classes, especially the misclassification of Early Blight and Late Blight that visually exhibited similar diseases. EfficientNetV2L achieves slightly higher results but nevertheless has a few comments about all categories of classification errors. EfficientNetV2S and ConvNextBase, however, show excellent diagonal matrices with low off-diagonal error that shows good but imperfect classification capacity. In general, the confusion matrix discussion justifies the fact that Lightweight ViT is incomparable in the terms of precision and classes separation and should be regarded as the most accurate model in terms of fine-grained disease detection.

- **Classification Report Analysis**

The classification reports provide further details of the precision, recall, and F1-score of each of the models in the three classes of disease. The Lightweight ViT performance on both the validation and test set shows a lightweight version of ViT with a perfect score (1.00) on each of the three performance metrics, per target class. This indicates that the model not only predicts but predicts with they utmost certainty and consistency in all groups and therefore there is a superior generalization. ConvNeXtBase and EfficientNetV2S come right behind, though, with macro-averaged F1-scores of 0.97 to 0.99 meaning effective class performance with a few occasional misclassifications. EfficientNetV2L has a reasonably high performance (~0.93 macro F1), but it makes a few mistakes in determining whether it is Early or Late Blight, which brings down its rankings. EfficientNetV2M, in its turn, perform very poorly, with macro and weighted F1-scores averaging at about 0.70-0.75, denoting its inability to differentiate between visually similar symptoms. Through this deep understanding, it is important to say that, even Lightweight ViT can not only be very accurate, but also guarantee uniform and under-class behavior, which is vital in sensitive agricultural diagnostics.

- **ROC Curve and AUC Analysis**

Receiver Operating Characteristic (ROC) curves, together with the so-called Area Under Curve (AUC) numbers, can give a graphical and numeric indication of the classification thresholds of the models. Having the highest AUC score of 1.00, which means that the sensitivity and specificity are 100 percent ideal at all threshold values, we consider the Lightweight ViT model as the most promising one. Its ROC curve sharply rises to the left upper corner, which means the maximum separation between the classes. ConvNeXtBase and efficiencyNet v2s also exhibit high AUC scores (0.98- 0.99) indicating high discrimination power, but do exhibit some overlap in some of the classes too. When it comes to efficientNetV2L, it performs pretty good, but then with a bit less AUC because of higher false positives and false negatives. EfficientNetV2M has a lower ROC curve, and it has the lowest AUC (~0.85-0.90) of all the models, confirming poorer class separation of the model.

Overall, the ROC and AUC methodology supports Lightweight ViT, not only because of its high quality, but because of its trustworthy reliability in confidence calibration and threshold-free superiority of performance, and therefore it would be the preferred tool in precision agriculture or in a fully automated disease detection system.

4.4 Summary

This chapter thoroughly tested the suggested Lightweight ViT model using four other models that may be considered state of the art models of deep CNN models using the same training setting. None of the other models could compete with the Lightweight ViT that gave a classification of perfect accuracy on the validation and the test dataset, with AUC scores of 100. Although it has the fewest parameters and model size (~1.3M parameters and 4.99 MB), it generalized better, trained very efficiently and was less prone to overfitting. Although EfficientNetV2S and ConvNeXtBase were also quite effective, they were quite larger and therefore not so appropriate to be deployed in an environment with limited resources. The lowest scores were obtained on all evaluation metrics with EfficientNetV2M. The Lightweight ViT could be viewed as the most accurate, lightweight, and handy fix by using confusion matrix analysis, classification reports, and ROC curves. These results confirm the possibility of using transformer-based architectures in the real-world, mobile, and at the edge in agricultural applications in lightweight classification tasks.

Chapter 5

Engineering Standards and Design Challenges

The chapter gives a clear explanation of the experimental results, as compared to those found on existing literature. It provides interpretation of the results, the strengths and the limitations of the proposed lightweight model and the applications of the mobile-based potato leaf disease detection in an agricultural field.

5.1 Compliance with the Standards

The suggested system is elaborated according to the existing engineering and technological requirements to achieve reliability, scalability, and safety of a user. There are established practices in the preparation of a recognized dataset in image processing and deep learning such as labeling the images properly, augmenting, and normalizing the images to keep the data unbiased in quality. Its design and training of the model follows widely used machine learning development standards, e.g. the evaluation metric of accuracy, precision, recall and F1-score are used to ensure the performance of the model is fairly measured. The mobile application element would be developed according to the common software engineering recommendations of usability, responsiveness, and accessibility, which would facilitate the application among varied users, including farmers with less knowledge of technology. The design also focuses on interoperability and device compatibility standards such that the solution is used in harmony with different smartphones without special hardware. The privacy of data and ethics are also upheld, and personally identifiable information is not captured and used in the analysis process of the images. These standards provide the system with reliable operation, customer confidence, and preparedness to utilize the system in the real world within agriculture scenarios.

5.1.1 Software Standards

The software development procedure of this system applies known recognized software engineering standards in terms of efficiency, maintainability purpose and reliability of the same. The mobile app is a creation based on platform-respective frameworks which satisfy Android app development protocols and are perfectly implemented to perform appropriately in all devices and operating system iterations. Coding obeys clean code, modular design and good documentation, hence can be modified and updated later with ease. The deep learning model integration follows generic model deployment conventions e.g. converting deepl models using optimized formats (e.g. TensorFlow Lite) to run quickly on mobile devices without sacrificing accuracy. Git is used to maintain a version of the modified files to trace the updates and collaborate. The user interface is designed in accordance with the basic principles of usability and accessibility that enables farmers and other agricultural workforce to use the application without any prior technical knowledge. Observing these software guidelines, the system provides a compromise between the technical strength, ease of use and future flexibility.

5.1.2 Hardware Standards

It is possible to compile the hardware specifications of this system since they are required to have an efficient functioning of not only the deep learning model but also the mobile application. Standard computing hardware, including processing power (e.g. GPUs) is utilized to meet the computational requirements of transformer-based models to train the model. In the deployment, mobile application execution will be multi-faceted to support their deployment on mid-range Android phones, more feasible to the farmers and the workers in the agricultural sector. Hardware standards are aimed at the optimization of performance with minimal energy usage making sure that the application will be stable and even does not demand the use of more advanced devices. The most important factors are compatibility with the camera modules that are readily available in the market on smartphone phones since, directly, the quality of images taken influences the accuracy of the model. The amount of storage and the use of the memory is relegated to a decent parameter to make sure that the application does not become slow and it does not overburden the device. Observing such hardware standards, the system will be viable, effective economically, and applicable in the field of agriculture where one might work with a limited supply of resources.

5.1.3 Communication Standards

Communication standards used in this system only aim at facilitating effective exchange of data between the server and mobile application when updating the model or when communicating remotely. Although this application can be used offline, it must have some form of internet connectivity due to the need to update the model to improve it, bug fixes, and addition of more data in the dataset. The transmission of data is secured by standard communication protocols like HTTPS so that personal information of a model and users cannot be encoded by others. The mobile application is also designed to consume low bandwidth so that updates and transfer of information can be carried out over even remote locations where the network coverage is not good. When necessary it used compression methods to narrow down the size of the images that are being transmitted and not to a great extent affect the quality. Through these communication specifications, the system will be reliable, secure and efficient and at the same time available to the end-users within different network setups.

5.2 Impact on Society, Environment and Sustainability

The potato leaf disease detection system that has been proposed can cause a positive change in agricultural and rural sectors. Being able to identify the disease early and accurately with a transformer model weighing just a few kilograms, farmers will be able to act in time thus reducing crop losses, which will result in enhanced productivity and food security. This minimizes the use of chemical pesticides, and consequently, it decreases the pollution of the environment and healthier farming agriculture. The mobile design makes the initiative available to farmers who live in far-away locations, enabling widespread farming to use technology without the purchase of costly gadgets. In terms of sustainability, the system promotes resource optimization, and does not result into wastage of water, fertilizer and pesticides. Long term, this will help lower the cost of production in agriculture, enhancing the lives of farmers, and supporting green agriculture which is among the objectives of sustainable agriculture and environment enshrined in the global agenda.

5.2.1 Impact on Life

The system has the potential to provide meaningful differences in day to day lives of the farmers and the agricultural workers. It saves time and cost as it eliminates the need to consult the experts and run tests in laboratories to detect the disease since it is a simple mobile program that detects the disease within the shortest time possible. This gives the farmers the opportunity to make more decisive and quicker decisions with regards to crop management and ultimately safeguard their income and decrease stress related to unforeseen losses. The lower permissible use of toxic substances also allows the farming families and communities around to live in healthier conducive conditions since the chances of exposure to toxic substances are minimal. Long-term benefits of this technology may facilitate knowledge transfer and digital literacy among the people in the rural areas and produce a self-taught and self- governed agricultural society.

5.2.2 Impact on Society & Environment

The system proposed here can be of benefit to society and the environment as it will lead to more sustainable agriculture. Treatment of croplands in the early stages of commercial plant establishment would help farmers curb the potential crop loss on a large scale and guarantee improved food security among the members of the society. The system allows reducing the unnecessary use of pesticides thus protecting soil quality, biodiversity, and avoiding other unwanted processes related to harmful chemical entry into surrounding water sources. This is not only beneficial to the ecosystem but also beneficial to the health of the population that is decreased by exposures to harmful substances. As well, the technology promotes the employment of intelligent farming algorithms that accelerate innovations and digitalization in the countryside. Long term, it helps create a stronger agricultural sector that would be adaptation to environmental problems whilst protecting livelihoods.

5.2.3 Ethical Aspects

Ethical aspects have their place during technical development and implementation of this system because it is a direct consequence of farming effect and decision-making. The technology should maximize precision and trustworthiness so that there is no misdiagnosis that can occur to treat the crop unnecessarily or lose overall yield. Preservation of information and data of farmers is also crucial, data on location and farming should be kept and used in an accountable manner where privacy and confidentiality are respected. Besides, the system must be made affordable and accessible whereby small-scale farmers should not be locked out of its provisions because of monetary or technical accessibility. The acceptance and familiarity of the model and its recommendations could be hampered due to lack of clarity on how the model is used or decisions are arrived at and transparency will remove the mistrust which some users would have to the model or the results. In resolving the dramatic elements, the system has the ability to promote equity, accessibility, and ethical innovation in agriculture.

5.2.4 Sustainability Plan

This system can only be maintained through constant development, use in a sustainable way, and easy availability to farmers over prolonged periods. Updates of the model with fresh, various data on potato leaf diseases would assist in keeping the model accurate as time goes by and as environmental factors and the way diseases develop alter. Lightweight transformer structures will make the solution energy efficient and suitable to run on the mobile devices that do not necessitate intensive hardware requirements hence lower the cost of operation as well as environmental impact. Awareness and training programs will be utilized to equip farmers on how to incorporate the system in their other farming habits and practice so that there is a continuous adoption. Partnership with agriculture groups, research centers, and local areas can extend the support and finance the technology to maintain it low-cost and accessible, making the technology appreciated by many. The system can be an effective and endurable means of preserving crop health and food security because of its mixture of technical efficacy and social responsibility.

5.3 Project Management and Financial Analysis

The successful implementation of the current research greatly relied on the ability to estimate the accurate cost of the project in terms of both effective project management and financial planning, especially because of the fact that the study was geared towards the development of lightweight transformer-based model and a mobile application to detect the potato leaf disease in real time. This method focused on resource effective utilization, cost-effectiveness, and feasibility to make it very practical to be inspected by farmers and other stakeholders of agriculture. In this section, the authors list the sequence of operations carried out during the project, as well as the assistant to the budget optimisation and sustainability.

5.3.1 Project Planning and Task Management

The study used project management approach that was agile-based in order to be flexible, iterative, and track progress. The actual work was split into well-bounded stages: acquiring a dataset, preprocessing, model design and training, evaluation, deployment, and mobile applications integration. The stages were characterized by certain objectives, deliverables and review points in order to maintain a steady progress.

The first step consisted of gathering publicly available potato leaf datasets and resizing, normalization, and augmentation of the data as a preprocessing step to increase quality and diversity of the data. The selection of transformer-based architectures, namely Scratch ViT and MobileViT, as well as their further model development were based on the balance between accuracy and computational efficiency. The hyperparameter tuning and training were done sequentially and hyperparameter configuration with the best performance was determined using evaluation results.

One of the significant steps was the deployment of models whereby the model trained was saved into a TensorFlow Lite format to be used in a mobile device. The mobile app then was shaped with Flutter technology, and the lightweight model is incorporated into classifying disease in real-time. Lastly, tough testing was conducted as a means of determining usability and compatibility among multiple devices.

There was version control and sharing of files with collaboration systems like GitHub and Google Drive. The tasks were delivered on time and adjusted in case of changes due to regular reviews of the milestones.

1. Resource and Time Allocation

The project was completed over approximately five months, with specific phases distributed as follows:

- **Month 1:** Dataset collection, literature review, requirement analysis

- **Month 2:** Data preprocessing (resizing, normalization, augmentation), initial model setup
- **Month 3:** Transformer architecture design, model training, hyperparameter tuning
- **Month 4:** Model evaluation, TFLite conversion, mobile app integration
- **Month 5:** Final testing, UI refinement, documentation, and reporting

Google Colab's GPU resources were leveraged for training, eliminating the need for local high-end hardware and accelerating experiments.

2. Financial Analysis

The project would also be cost-efficient as it was based on open-source frameworks and rather free cloud resources, as well as using publicly available datasets. There was little spending and thus mostly on the very basic infrastructure like the availability of the internet connectivity.

3. Scalability and Future Investment

The modular design of the model and mobile application ensures future scalability. With minimal additional cost, the framework can be extended to include more crop types, additional disease classes, or offline processing features for low-connectivity environments. The system's lightweight design makes it highly deployable in resource-constrained agricultural regions, creating significant potential for real-world adoption at scale.

The phased project structure followed these milestones:

- Requirement Analysis and System Design – Defining objectives, tools (TensorFlow, Flutter, Colab), and architecture.
- Dataset Preparation and Preprocessing – Collecting and preparing potato leaf datasets with resizing, augmentation, and normalization.
- Model Development and Training – Training transformer-based models (ViT, MobileViT) using GPU resources.
- Model Deployment – Converting trained model into TFLite and ensuring mobile compatibility.
- Mobile Application Development – Building a Flutter-based Android application and integrating the model.
- Documentation and Reporting – Preparing thesis, research outcomes, and final reporting.

5.3.2 Tools and Platforms Used

Table 5.1 provides an overview of the essential tools and platforms used for model training, development, deployment, and version control throughout the project.

Table 5.1: Details of tools and platform used

Category	Tool/Platform	Purpose
Model Training	Google Colab (GPU)	Training models
Programming Language	Python	Core ML development and preprocessing
DL Framework	TensorFlow / Keras	Model building, training, and TFLite conversion
Mobile Development	Flutter	Cross-platform mobile app development
IDE	Android Studio	Testing and packaging the Android app
Version Control	Git / GitHub	Code versioning and collaboration

5.3.3 Financial Analysis

This project was designed to be cost-effective and accessible for students or independent researchers. Below is an approximate cost breakdown:

Table 5.2: Estimated Cost and Financial Analysis.

Resource/Item	**Estimated Cost (BDT)**	**Remarks**
Google Colab Pro (optional)	3500	Faster training with extended GPU time
Personal Computer/Laptop	Existing Resource	Used for mobile app development and testing
Internet Access	4200	Required for training and deployment
Flutter & Android Studio	0	Open-source development tools
Cloud Storage/Backup (optional)	300	Google Drive or similar services
Total	8000 BDT	–

5.4 Complex Engineering Problem

Engineering The process of creating a lightweight transformer-based detection system of the potato leaf disease deals with a complex engineering problem which consists of the combination of machine learning, mobile application development, and agricultural requirements. The problem is designing a model which is both very powerful and computationally accurate that it can be used within low power mobile machines in rural regions. This involves a tradeoff between data preprocessing, model architecture and optimization approaches to deal with both lighting, image quality and environmental variability whilst maintaining a minimal overhead on resources used. Besides, the inclusion of the model in a mobile interface that is easy to use presents design limitations that are associated with usability, offline access, and data confidentiality. Satisfying these constraints requires a strong knowledge of computer vision and limitations of hardware, as well as understanding users behavior, which makes the problem interdisciplinary and highly technical.

5.4.1 Complex Problem Solving

To successfully complete this research project, several aspects of complex engineering problem-solving were involved—ranging from algorithm selection and model integration to mobile optimization and real-time deployment. The following mapping demonstrates how the work aligns with the Engineering Problem (EP) framework (see Table 5.3):

Table 5.3: Mapping with Complex Engineering Problem.

EP1	EP2	EP3	EP4	EP5	EP6	EP7
Depth of Knowledge	Conflicting Requirements	Depth of Analysis	Familiarity of Issues	Applicable Codes	Stakeholder Involvement	Interdependence
✓	✓	✓	✓		✓	✓

Justifications:

- EP1 – Depth of Knowledge: The research demands high level knowledge of computer vision, transformer architectures and diagnosis of agricultural diseases. It uses deep learning skills integrated with a plant pathology understanding to develop a light and precise detection scheme that can be deployed in real-time and in a mobile environment.
- EP2 – Range of Conflicting Requirements: Ensuring accuracy-based detection and low demands on the mobile device in terms of computation power provide mutually conflicting bottlenecks. The system has to be efficient in the rural areas where internet and hardware infrastructure is lower.
- EP3 – Depth of Analysis: The study entails extended data preprocessing, the comparison of the model architecture (MobileViT vs. ViT from scratch), and an assessment to different conditions in the environment with the

environmental images.

- EP4 – Familiarity of Issues: Deals with such familiar problems of agriculture such as misidentification of the disease and the absence of real-time tools and presents AI-powered solutions adjusted to field conditions.
- EP6 – Level of Stakeholder engagement: The project focuses on stakeholders who are involved as the key stakeholders, namely farmers, agricultural extension workers, and researchers, hence its usability and its practical field desirability.
- EP7 – Interdependence: Comprises multidomain integration: AI, mobile computing, agriculture, and user interface design, each of them with an impact on the overall system performance and adoption.

Mapping with Knowledge Profile

Table 5.4: Mapping with knowledge Profile.

K3	K4	K5	K6	K8
Engineering Fundamentals	Specialist Knowledge	Engineering Design	Engineering Practice	Research Literature
✓	✓	✓	✓	✓

Justifications:

- K3 – Engineering Fundamentals: The work is based on deep learning and computer vision fundamentals used to solve the agricultural problems.
- K4 – Specialist Knowledge: Uses specialized techniques of transformer structure and optimization approach to carry out mobile vision problems.
- K5 – Engineering Design: Designs a comprehensive detection system, such as choosing an architecture, preparing a dataset, and a deployment strategy.
- K6 – Engineering Practice: Tests and implements the system within the real-world restrictions, i.e., mobile device limitations and offline capability.
- K8 – Research Literature: constitutes an extension of literature review conclusions and makes mention of identified research gaps.

5.4.2 Engineering Activities

The development lifecycle of the project involved complex engineering activities across multiple domains, including AI model training, evaluation, software integration, user interface development, and mobile deployment. These activities reflect the real-world complexity of delivering a working healthcare AI system under practical constraints (see Table 5.5).

Table 5.5: Mapping with Complex Engineering Activities.

EA1	EA2	EA3	EA4	EA5
Range of Resources	Level of Interaction	Innovation	Societal & Environmental Consequences	Familiarity
✓	✓	✓	✓	✓

Justification:

- EA1 – Range of Resources: Employs cloud GPU resources to train, uses open source data-sets of potato leaf disease images, and uses mobile hardware to deploy.
- EA2 – Level of interaction: Shall involve co-operation between AI developers and agriculturalists and end users (farmers and extension workers).
- EA3 – Innovation: It integrates mobile application deployment with a type of lightweight transformers that are uncommon in the current plant disease detection literature.
- EA4 – Social and Environmental impacts: Makes crop loss minimal and makes farming sustainable since diseases can easily be detected before it is too late, which can enhance food security.
- EA5 – Familiarity: The solution solves a familiar agricultural problem using a new AI technique and integration with mobile devices to increase accessibility and usability.

5.5 Summary

To conclude, the construction and creation of the transformer-based potato leaf disease detector became the project needing a thorough consideration of technical and practical aspects. This was done by adherence to pertinent engineering standards that guarantee accuracy, efficiency, and safety of users, as well as considering the real-life problems like shortage of hardware, dynamically changing images and a convenient mobile interface. There was also major emphasis on the balancing of the model performance against the limitations of the resources so that the system would be able to provide quality results throughout the rural and agricultural environment. In this chapter, the engineering principles, ethical and usability objectives were described on how they guided such design choices so that the resulting solution would be not only innovative but also practical to implement it in sustainable agricultural practices.

Chapter 6

Conclusion

This chapter outlines the contributions to the research, the findings and the general result of the study. It also proposes future possible directions in the continued expansion of the model to other crops, further advanced architecture of lightweight transformers, or simplifying the mobile application toward wider adoption.

6.1 Summary

The proposed research was based on diagnosing diseases of potato leaves by using a lightweight transformer architecture, coupled with the detection of such diseases using mobile applications in real-life situations. The research set out to solve the shortcomings of the general plant disease diagnostic strategies that are time consuming, demand professional and skilled expertise and are unavailable to rural farmers. In an attempt to circumvent these difficulties, an entire workflow was established, including data collection and preprocessing, model development, evaluation and deployment.

To achieve consistency and therefore a better picture performance the dataset comprising of potato leaf images was prepared by resizing, augmentation, and normalization. The study compared two transformer-based models, Vision Transformer (ViT) and MobileViT based on both accuracy and computational efficiency. The lightweight nature of MobileViT made it very suitable to be deployed to mobile devices without compromising on the competitive levels of accuracy.

The system was also incorporated into a mobile app in which accuracy was as fast as the user snapped or uploaded a leaf image and obtained immediate detection. Such a mix of deep learning and mobile technology can help the farmers detect disease early to allow them to take action and minimize losses. Due to the nature of transmission, the results prove that lightweight transformer architectures are a feasible and pragmatic system of detection of plant diseases in agricultural practical application.

6.2 Limitation

This study of identifying the potato leaf disease through the lightweight transformer architecture produced some good results, but some limitations still persist. The data provided are rather diverse, yet it was gathered on the basis of sources available publicly, and thus it might not fully accommodate all conditions possible in practice, including radical differences in lighting, soil backgrounds, or the presence of rare stages of the disease. Consequently, the performance of the model may be compromised when it is used in the field that was not seen during training.

The quality and consistency of the images is another factor that is limited. Part of the images in the dataset had been of different resolutions, focus, and clarity, a fact that may affect the feature extraction of the model. Also, although data augmentation methods were used to enhance generalization, they might not exactly resemble the natural variation that would be present in real-life farming situations.

The model has been mostly simulated in purpose driven computing environments, i.e., there is little validation in terms of network latency, hardware limits on the devices as well as the user engagement in the mobile app. In addition, despite the decision to adopt the lightweight transformer architecture based on efficiency, the number of trade-offs between minimizing the cost of computation and characterizing highly complex disease trends remains.

Lastly, this experiment dwelled on leaf level detection of disease. It did not take on whole-plant analysis, the conditions of the soil, and any environment conditions that could affect the health of the plants as well. Future research to help deal with these limitations can lead to a stronger, more flexible, and farm-ready system that farmers and agricultural workers can use.

6.3 Future Work

Although the proposed work proves the efficiency of the lightweight transformer architectures in the identification of potato leaf diseases, there remains a number of things that can be improved during the future research. Among possible directions, the extension of a more diverse and larger data set that includes the variety of potato, seasonal changes, and environmental aspects is possible. This would aid the model to learn more durable features and enhance its optimization to the real world.

The other significant extension is enhancement with explainable AI (XAI) in the mobile application. It would help the farmers and the agricultural experts to not only know the disease prediction but also get to know about the areas of the leaves that affected the model decision. These levels can promote trust and foster improved decision making in the discipline.

It is possible to also improve the current mobile app to operate in complete offline mode so that it could be used in the areas with low or even no internet connectivity. It could become a more complete tool in farming with the ability to add such features as estimating the level of diseases, prescribing treatment, and creating automated reports to agricultural agencies.

Last but not least, combining lightweight transformer models with other deep learning methods that are efficient e.g. pruning, quantization, or knowledge distillation can be a future work direction. This would also cut down on required computation making the system more friendly to low-resource devices without compromising on high accuracy. Working on these aspects, the proposed system will develop into a sustainable earliest disease detection and intelligent agricultural system that is easy to scale and practice.

References

[1] Hughes, D., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060.

[2] Shi, Y., Han, L., Kleerekoper, A., Chang, S., & Hu, T. A Novel CropdocNet for Automated Potato Late Blight Disease Detection from the Unmanned Aerial Vehicle-based Hyperspectral Imagery. arXiv 2021. arXiv preprint arXiv:2107.13277.

[3] Meng, Q., Guo, J., Zhang, H., Zhou, Y., & Zhang, X. (2025). A dual-branch model combining convolution and vision transformer for crop disease classification. PloS one, 20(4), e0321753.

[4] Hu, X., Jiang, H., Liu, Z., Gao, M., Liu, G., Tian, S., & Zeng, F. (2025). The Global Potato-Processing Industry: A Review of Production, Products, Quality and Sustainability. Foods, 14(10), 1758.

[5] Lamichhane, S., Neupane, S., Timsina, S., Chapagain, B., Paudel, P. P., & Rimal, A. Potato Late Blight Caused by Phytophthora Infestans; an Overview on Pathology, Integrated Disease Management Approaches, and Forecasting Models. Integrated Disease Management Approaches, and Forecasting Models.

[6] Zitter, T. A., & Loria, R. (1986). Detection of potato tuber diseases and defects.

[7] Ghosh, A., Ghosh, T. K., Das, S., Ray, H., Mohapatra, D., Modhera, B., ... & Bandyopadhyay, R. (2022). Development of electronic nose for early spoilage detection of potato and onion during post-harvest storage. Materials NanoScience, 9(2), 101-114.

[8] Kulkarni, P., Karwande, A., Kolhe, T., Kamble, S., Joshi, A., & Wyawahare, M. (2021). Plant disease detection using image processing and machine learning. arXiv preprint arXiv:2106.10698.

[9] Mehta, S., & Rastegari, M. (2021). Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. arXiv preprint arXiv:2110.02178.

[10] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[11] Oteyo, I. N., Marra, M., Kimani, S., Meuter, W. D., & Boix, E. G. (2021). A survey on mobile applications for smart agriculture: Making use of mobile software in modern farming. SN computer science, 2(4), 293.

[12] Paul, H., Udayangani, H., Umesha, K., Lankasena, N., Liyanage, C., & Thambugala, K. (2025). Maize leaf disease detection using convolutional neural network: a mobile application based on pre-trained VGG16 architecture. New Zealand Journal of Crop and Horticultural Science, 53(2), 367-383.

[13]Shaheed, K., Qureshi, I., Abbas, F., Jabbar, S., Abbas, Q., Ahmad, H., & Sajid, M. Z. (2023). EfficientRMT-Net—an efficient ResNet-50 and vision transformers approach for classifying potato plant leaf diseases. Sensors, 23(23), 9516.

[14]Adhikari, S. (2024). Advancements in Agricultural Technology: Vision Transformer-Based Potato Leaf Disease Classification. Journal of Soft Computing Paradigm, 6(2), 169-185.

[15]Reis, H. C., & Turk, V. (2024). Potato leaf disease detection with a novel deep learning model based on depthwise separable convolution and transformer networks. Engineering Applications of Artificial Intelligence, 133, 108307.

[16]Arshad, F., Mateen, M., Hayat, S., Wardah, M., Al-Huda, Z., Gu, Y. H., & Al-antari, M. A. (2023). PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction. Alexandria Engineering Journal, 78, 406-418.

[17]Sinamenye, J. H., Chatterjee, A., & Shrestha, R. (2025). Potato plant disease detection: leveraging hybrid deep learning models. BMC Plant Biology, 25, 647.

[18]Li, G., Wang, Y., Zhao, Q., Yuan, P., & Chang, B. (2023). PMVT: a lightweight vision transformer for plant disease identification on mobile devices. Frontiers in Plant Science, 14, 1256773.

[19]Thai, H. T., & Le, K. H. (2025). MobileH-Transformer: Enabling real-time leaf disease detection using hybrid deep learning approach for smart agriculture. Crop Protection, 189, 107002.

[20]Pacal, I., & Işık, G. (2025). Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification. Neural Computing and Applications, 37(4), 2479-2496.

[21]Zhao, Y., Li, Y., Wu, N., & Xu, X. (2024). Neural network based on convolution and self-attention fusion mechanism for plant leaves disease recognition. Crop Protection, 180, 106637.

[22]Jahangir, R., Sakib, T., Baki, R., & Hossain, M. M. (2023, November). A Comparative Analysis of Potato Leaf Disease Classification with Big Transfer (BiT) and Vision Transformer (ViT) Models. In 2023 IEEE 9th International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE) (pp. 58-63). IEEE.

[23]Bhowmik, A. C., Ahad, M. T., Emon, Y. R., Ahmed, F., Song, B., & Li, Y. (2024). A customised Vision Transformer for accurate detection and classification of Java Plum leaf disease. Smart Agricultural Technology, 8, 100500.

[24]Kamal, S., Sharma, P., Gupta, P. K., Siddiqui, M. K., Singh, A., & Dutt, A. (2025). DVTXAI: a novel deep vision transformer with an explainable AI-based framework and its application in agriculture. The Journal of Supercomputing, 81(1), 1-32.

[25]Bajpai, A., Sahu, S., Tiwari, N., Srivastava, V., & Yadav, S. (2024, August). An efficient approach for potato leaf disease classification using cascaded cnn-

transformers. In 2024 IEEE 12th International Conference on Intelligent Systems (IS) (pp. 1-6). IEEE.

[26]Jahangir, R., Sakib, T., Baki, R., & Hossain, M. M. (2023, November). A Comparative Analysis of Potato Leaf Disease Classification with Big Transfer (BiT) and Vision Transformer (ViT) Models. In 2023 IEEE 9th International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE) (pp. 58-63). IEEE.

[27]Meng, Q., Guo, J., Zhang, H., Zhou, Y., & Zhang, X. (2025). A dual-branch model combining convolution and vision transformer for crop disease classification. PloS one, 20(4), e0321753.

213-15-4605

ORIGINALITY REPORT

14%	10%	8%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	www.mdpi.com Internet Source	1%
4	Submitted to United International University Student Paper	<1%
5	arxiv.org Internet Source	<1%
6	ikee.lib.auth.gr Internet Source	<1%
7	Sushil Kamboj, Pardeep Singh Tiwana. "Innovations in Computing", CRC Press, 2025 Publication	<1%
8	Shankar Babu, Mahesh Babu Kota. "Synergies in Smart and Virtual Systems using computational intelligence", CRC Press, 2025 Publication	<1%
9	Submitted to University of Carthage Student Paper	<1%
10	Submitted to Lincoln University Student Paper	<1%
11	avestia.com Internet Source	<1%

12	Qingduan Meng, Jiadong Guo, Hui Zhang, Yaoqi Zhou, Xiaoling Zhang. "A dual-branch model combining convolution and vision transformer for crop disease classification", PLOS ONE, 2025 Publication	<1%
13	medium.com Internet Source	<1%
14	Submitted to University of Ulster Student Paper	<1%
15	www.cubert-hyperspectral.com Internet Source	<1%
16	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	<1%
17	svu-naac.somaiya.edu Internet Source	<1%
18	bdta.abcd.usp.br Internet Source	<1%
19	Egling, Theodore. "Differential Evolution Optimization Algorithms and its Application in Machine Learning Based Disease Detection", University of South Africa (South Africa) Publication	<1%
20	globalpresshub.com Internet Source	<1%
21	O. P. Verma, Seema Verma, Thinagaran Perumal. "Advancement of Intelligent Computational Methods and Technologies - Proceedings of I International Conference on Advancement of Intelligent Computational	<1%

Methods and Technologies (AICMT2023)", CRC Press, 2024
Publication

22	thesai.org Internet Source	<1 %
23	www.frontiersin.org Internet Source	<1 %
24	export.arxiv.org Internet Source	<1 %
25	Asif Raza, Abdul Hammed Pitafi, M. Kahsif Shaikh, Khaliq Ahmed. "Optimizing potato leaf disease recognition: Insights DENSE-NET-121 and Gaussian elimination filter fusion", Heliyon, 2025 Publication	<1 %
26	Burak Gülmez. "A Comprehensive Review of ConvolutionalNeuralNetworksbased Disease Detection Strategies in Potato Agriculture", Potato Research, 2024 Publication	<1 %
27	Submitted to Middlesex University Student Paper	<1 %
28	Satyajit Chakrabarti, Ashiq A. Sakib, Souti Chattopadhyay, Sanghamitra Poddar, Anupam Bhattacharya, Malay Gangopadhyaya. "Interdisciplinary Research in Technology and Management", CRC Press, 2024 Publication	<1 %
29	dokumen.pub Internet Source	<1 %
30	journals.nasspublishing.com Internet Source	