A Comparative Study of ConvNeXt-Tiny and Swin-Tiny for Multi-Label Plant Disease Classification on Real-World Leaf Images

<u>Submitted By:</u>

MD. KHALID HASAN RIAD

221-35-911

Department of Software Engineering

Daffodil International University

<u>Supervised By:</u>

Ms. Tapushe Rabaya Toma

Assistant Professor

Department of Software Engineering

Daffodil International University

Thesis submitted in fulfillment of the requirements for the award of the degree of Bachelor of Science in Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

Fall 2025

APPROVAL

This thesis titled on "**A Comparative Study of ConvNeXt-Tiny and Swin-Tiny for Multi-Label Plant Disease Classification on Real-World Leaf Images**", submitted by **MD. Khalid Hasan Riad (ID: 221-35-911)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. S. M. Hasan Mahmud
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

13.12.25

Internal Examiner 1

A.H.M Shahariar Parvez
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Tapushe Rabaya Toma
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 3

Khalid Been md. Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Dr. Md Sazzadur Rahman
Professor
Institute of Information technology
Jahangirnagar University, Bangladesh

SUPERVISOR'S DECLARATION

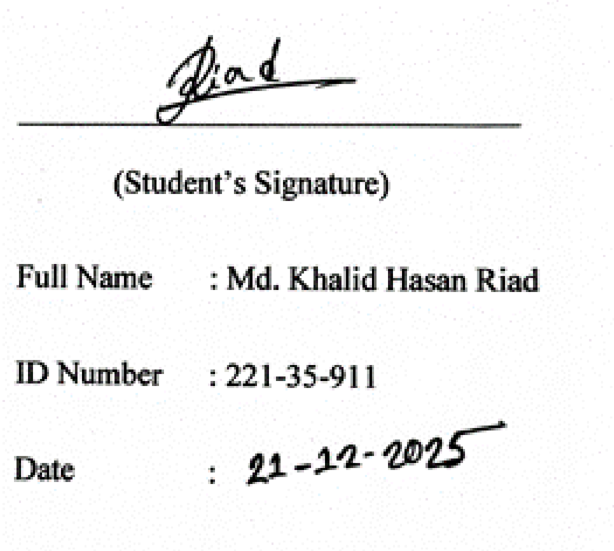
I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.

(Supervisor's Signature)

Full Name : Tapushe Rabaya Toma
Position : Assistant Professor
Date : 21.12.25

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.



Riad

(Student's Signature)

Full Name : Md. Khalid Hasan Riad

ID Number : 221-35-911

Date : 21-12-2025

ACKNOWLEDGEMENTS

I want to show my deep gratitude to my supervisor because he provided essential backing and skilled direction and enduring tolerance which helped me complete this work. Their absolute faith in cryptocurrency volatility understanding led them to accept multiple testing and revision processes.

I am also sincerely grateful to the larger open-source and scholarly communities. The research uses tools and datasets and intellectual work from developers and academics who have contributed to Python and pandas and scikit-learn and time-series modeling libraries.

I want to show my deepest gratitude to everyone who supports me including my family members and my friends and my work colleagues. Their feedback on my first drafts together with their research presentation and continuous assistance helped me understand the thesis better while making it more effective.

ABSTRACT

The current deep learning systems which detect plant diseases in farms require substantial financial resources because they operate with lab-based data and single-label classification methods which fail to function in actual field conditions where multiple diseases can affect one leaf. This study uses the Plant Pathology 2021 dataset, which has field images, to test the ConvNeXt-Tiny and Swin-Tiny architectures for classifying apple diseases. Both models started with pretrained weights from ImageNet 22k. The model used class-weighted loss to address the significant class imbalance problem. The evaluation process measured how long it took for computers to generate predictions and also checked the accuracy of these predictions. The results indicate that both models produce successful results. Swin-Tiny is better at remembering things and making guesses, while ConvNeXt-Tiny is better at both. The two models successfully detect diseases which present different symptoms but they fail to identify multiple diseases that exist simultaneously. The results show that the architecture you choose should depend on what you need to do with it. The two models offer different benefits for application development because ConvNeXt achieves better accuracy through its quick training time but Swin provides better real-time decision performance and memory efficiency. This study provides real-world evidence to help make architectural decisions in agricultural AI by showing that multi-label frameworks better show how diseases happen together in real life.

Table of Contents

List of Figures

List of Tables

CHAPTER 1 INTRODUCTION

1.1 Introduction

Food security throughout the world still places the most significance on agriculture. It serves over 2.5 billion directly. Plant diseases, however, are a major issue to the farming sector since they are able to reduce the crop yield by up to 40 percent annually on the global scene. This is more expensive than 220 billion to the economy (Savary, et al., 2019). Many of the conventional methods of detecting plant disease rely on the examination of plants by hand using trained agronomists. In the developing countries, smallholder farmers lack time and money to do so. In precision agriculture, artificial intelligence specifically deep learning is considered a transformative technology because it is highly needed to identify diseases with high precision, rapidity and scalability.

Automated plant disease classification has been very successful with the recent advancements in computer vision. Convolutional neural networks (CNNs) have been capable of achieving greater than 99 percent precision on regulated datasets such as PlantVillage that includes over 54,000 images of healthy and diseased plant leaves captured under the identical laboratory environment (Mohanty, Hughes, & Salathé, 2016). But such models are not always effective with real life farming scenarios where lots of diseases are occurring simultaneously, there is a complex background, the sun is moving and there is an obstacle in the way. Furthermore, the majority of the studies that are already researched are on single-label classification, which is too straightforward since it presupposes that each leaf node displays only one condition. This does not put in consideration that numerous diseases occur concomitantly.

These crucial gaps are filled in by this thesis by comparing two state-of-the-art architectures, ConvNeXt-Tiny and Swin-Tiny, systematically to classify multi-label plant disease on the Plant Pathology 2021 FGVC8 dataset, which consists of real-world field images of apple leaves. It is an excellent benchmark to operate the models that are going to be deployed since it has real-life problems, including non-uniform backgrounds, non-uniform camera angles, and combinations of diseases that occur naturally. The research offers a practical guideline on the choice of architecture that is most appropriate to use in the context of deploying agriculture by focusing on accuracy and computational efficiency.

1.2 Background and Motivation

By 2050, the world food demand is estimated to have grown by half because of population growth and dietary changes, and this will require a tremendous increment in the productivity of agriculture (Mottet, et al., 2017). One of the biggest impediments towards this aim is plant diseases. Diseases like scab, rust, and powdery mildew cost producers huge economic damages in the production of apples alone and intensive use of fungicides increases environmental and economic issues (Jiang, Chen, He, & Liang, 2020). To provide targeted intervention, early and accurate detection of diseases is required, which means that precise treatment that consumes fewer chemicals and that yields more will be provided and has less impact on the environment.

With the advent of deep learning, computer vision has been transformed and now feature extraction can be performed automatically, making it better than the traditional handcrafted feature extraction. ImageNet transfer learning has been especially useful in agricultural fields, where models trained on millions of general images can be used to solve more specific plant disease detection problems with relatively little labeled data (Müller, Jaeger, & Hanewinkel, 2019). With the advent of deep learning, computer vision has been transformed and now feature extraction can be performed automatically, making it better than the traditional handcrafted feature extraction. ImageNet transfer learning has been especially useful in agricultural fields, where models trained on millions of general images can be used to solve more specific plant disease detection problems with relatively little labeled data (Dosovitskiy, et al., 2020). The Swin Transformer proposed shifted window-based attention based hierarchical feature representation, which overcomes the computational drawbacks of conventional ViTs and still allows modeling complex visual representations (Liu, et al., 2021). Meanwhile, ConvNeXt showed that modernized CNNs based on transformer-like design features (larger convolutional kernels, LayerNorm, depth-wise convolutions) could perform the same as transformers at the same level of computational efficiency and induction biases as convolutions (Liu, et al., 2022). This architectural abduction poses fundamental questions concerning the design philosophy that would be more appropriate to domain specific applications like agricultural disease detection.

Though these architectural improvements exist, there is still a critical gap between model behaviour on curated data and its performance in farms. The models that are trained on the lab data, such as PlantVillage, tend to be much less accurate once applied to images of the field because of domain shift and drop in accuracy, which can make this model practically unusable (Maree, Fox, Al Rwahnih, Boonham, & Candresse, 2018). Also, the fact that the most common classification is based on single label does not take into consideration the biological fact that plants tend to have more than one infection happening at a time. Multi-label classification systems which are capable of identifying co-occurring illnesses can offer a more detailed diagnosis so that farmers can make sound treatment choices. These are the reasons why there is a necessity of the strict assessment of current architectures on real-world, multi-label agricultural data.

1.3 Problem Statements

Although the recent breakthroughs in the field of deep learning-based plant disease detection have come a long way, there are still a number of critical limitations which impede the move between research prototypes and real-life applications in the agricultural sector. According to this thesis, four major problems are outlined and discussed:

Problem 1: Dataset Realism Gap

Majority of the available studies on classification of plant diseases are based on controlled data sets that use laboratory set ups having uniform backgrounds, equal light and standardized camera distances. PlantVillage dataset is largely used in benchmarking but includes photos to plain backgrounds, which does not reflect the practice field conditions, resulting in models that do not generalize when applied in the real agricultural environment (Mohanty, Hughes, & Salathé, 2016). In the real world, the condition of light, partial occlusive by other plants, leaf maturity, and the backgrounds are dynamic and subject to change, and they pose a challenge to the robustness of models. This dataset bias generates a non-realistic capacity of performance that cannot be operationalized into a realistic utility.

Problem 2: Limitations in Single-Label Classification

Most systems of plant disease detectors use the single-label classification method, which presupposes that a single plant leaf can only have one condition (healthy or a single disease). It should be noted, however, that, in natural farming settings, it is likely that there are often two or more pathogens which infect the same plant tissue, especially when the plant is stressed or untreated. Single-label methods require models to choose the most pronounced visible symptom, and disregard the secondary infections that might need other treatment plans. This simplification renders the automated diagnostic systems less clinically relevant and their usefulness restricted to those farmers who require in-depth disease profiles.

Problem 3: Limited Architectural Comparisons on Agricultural Data

Although ConvNeXt and Swin Transformer are conflicting design philosophies (modernized CNNs vs. hierarchical transformers), specific agricultural applications between the two architectures are few. Comparative studies that exist are centered mainly on general computer vision benchmarks such as ImageNet, and both architectures are equally good in performance. Nevertheless, domain-specific features of agricultural images, including fine-grained texture structures, color changes that signify disease, and multi-scale manifestations of the symptoms might prefer one architecture to another (Liu, et al., 2022). It is crucial to know what design principles are more useful with regard to plant pathology imagery to inform the development of models in the future.

Problem 4: Accuracy-Efficiency Trade-offs

The systems of agricultural AI will have to weigh the accuracy of the classification with the efficiency of the underlying implementation, especially when such systems are to be deployed to resource-constrained edge devices or mobile platforms that can be carried by farmers. Although large models can have slightly better accuracy, they can be computationally intensive enough to be impractical in field diagnosis in real-time. This requires assessment models which take into account the predictive performance as well as the inference efficiency to determine the right architecture to deploy in various situations.

1.4 Research Objectives and Questions

The The main goal of the study is to provide the systematic comparative analysis of ConvNeXt-Tiny and Swin-Tiny models in multi-label plant disease classification on field images of a real-world setting and evaluate the predictive performance and energy consumption of the models to make the architectural decisions which can be applied in practice in agricultural practice. The thesis will have five aims:

- Dataset Preparation and Multi-Label Transformation: Reformat the plant pathology 2021 FGVC8 dataset to a 6-class multi-label model that allows separate detection of the presence of another disease as well as healthy, complex, frog eye leaf spot, powdery mildew, rust, and scab.
- Model Training and Optimization: Train both ConvNeXt-Tiny and Swin-Tiny with the same training settings that include the ImageNet-22k pretrained weights, class-weighted loss functions to mitigate label bias and a standardized data augmentation strategy to compare the models on equal grounds.

- Comprehensive Performance Evaluation: Compare the performance of the two models on a held-out test set with multi-label classification metrics, such as macro F1-score, micro F1-score, exact match ratio, hamming loss, and per-disease precision, recall, and F1-scores to get finer performance information.
- Computational Efficiency Analysis: Measure the computational properties of the two architectures, such as inference time, model size, number of parameters and the amount of memory needed by the GPU to determine the complexity of each architecture and its deployment capability on a limited resource platform.
- Architectural Insight Generation: Evaluate the performance disparities to determine which architectural components - convolutional inductive biases vs. self-attention mechanisms, hierarchical vs. uniform processing, local vs. global receptive fields are most useful in multi label plant disease classification.

This research addresses four key questions:

- RQ1: Which architecture (ConvNeXt-Tiny or Swin-Tiny) demonstrates better results in terms of multi-label classification on real-life apple disease images when measured based on macro F1-score as the main metric?
- RQ2: What is the performance of ConvNeXt-Tiny and Swin-Tiny in individual disease detection, specifically, when it comes to rare classes, such as powdery mildew, and more complicated ones, such as the complex disease category?
- RQ3: How do ConvNeXt-Tiny and Swin-Tiny differ in computational trade-offs based on the inference speed, model size, and memory costs of deploying the models in agricultural settings?

- RQ4: What are the differences in the attention processes of ConvNeXt-Tiny and Swin-Tiny in the detection of plant diseases, which can be observed using Grad-CAM visualization?

1.5 Contributions

This study contributes to the development of deep learning and agricultural artificial intelligence in a number of ways.

- Contribution 1: The plant disease identification system with multi labels in the real world. This paper describes a multi-label classification walk through that has been specifically created to address realistic multi-plant disease conditions where multiple infections occur simultaneously. This work allows making more clinically relevant predictions that are more representative of the actual agricultural conditions by transforming the original Plant Pathology 2021 dataset into independent binary classifications of each disease. The methodology offers a reproducible framework of future multi-label multi-agent-based agricultural classification activities such as data preparation, calculation of class weights, and multi-hot encoding.
- Contribution 2: Farm architectural comparison of data. The present research provides the first methodical comparison of two architectures ConvNeXt and Swin Transformer to classify plant diseases, with a similar set of conditions on field-realistic images. This experiment demonstrates domain-specific and architecture-specific performance properties of domain-specific problems, such as fine-grained recognition of disease symptoms, control of complex backgrounds, and multi-label

prediction accuracy, compared to comparisons of existing ones using general vision benchmarks.

- Contribution 3: A critical analysis that does not focus on the accuracy only. This study analyses predictive performance as well as computational efficiency where practical consideration of deployment is looked at which is often not considered in scholarly research. Measures such as speed of inference, memory footprint, and performance per disease assist practitioners to select appropriate models to be used in various agricultural scenarios, be it processing systems based on clouds or executing them on mobile devices at the edge.
- Contribution 4: Using open source and making things reproducible. All the code, trained models, and experimental set ups are made accessible to the public. This enables re-dobing of the results and assists in the future research. The implementation demonstrates how it is possible to conduct multi-label classification in the most appropriate way, such as splitting the dataset into groups, using class weighted loss functions, and recording all metrics. This can assist researchers that are developing AI systems in agriculture.

1.6 Summary

The necessity of having automated, accurate and efficient classification systems and the crucial issue of plant diseases discovery in contemporary farming were discussed in the present chapter to a significant extent. Plant diseases claim numerous crops across the globe and hence it is significant to detect them at an early stage to ensure the continuity of farming in the long term. Deep learning has performed well in controlled environments although the gap between the performance of the technology in the laboratory and in the real world remains large. In the case of the best design principles to apply to agriculture, the emergence of other types of architectural models such as ConvNeXt and hierarchical transformers such as Swin presents both opportunities and challenges. Most of the recent research does not recognize multi-disease co-occurrence, which is so common in field conditions, and instead makes use of single labels classification based on laboratory datasets. In order to fill these gaps, this thesis strictly compares two models, ConvNeXt-Tiny and Swin-Tiny, on the multi-label Plant Pathology 2021 FGVC8 dataset.

The research goals will focus on overall comparison of the two architectures in the areas of predictive accuracy and computational efficiency, which will give practical advice to apply AI in agriculture. The study is directed by four research questions, namely: what architecture has better overall performance, what are the differences in their performance on specific disease detection, what are their computational trade-offs and are their error patterns indicative of complementary performance?

The thesis provides a full multi-label classification system, the original direct ConvNeXt vs. Swin comparison of agricultural data, evaluation procedure, where deployment constraints are considered, and open-source implementation that can be used to reproduce it.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

This chapter investigates existing studies on deep learning-based classification of plant diseases, with a primary focus on architectural development, the generation of dataset, and multi-label classification systems. The review examines the evolution of computer vision, the traditional and modern deep learning. It discusses significance of CNN in agriculture, emergence of Vision Transformer and hierarchical counterparts such as Swin Transformer, and the resurrection of updated CNNs such as ConvNeXt. It further examines the evolution of datasets between the controlled lab images and the real-world conditions as well as the comparison of CNN and Transformer architectures as compared in other research. Research gap section determines the weaknesses that prevail in the current literature to be addressed by this study.

2.2 Literature Review

The evolution of using deep learning in forecasting plant diseases has changed significantly within the last ten years. Initial applications of Krizhevsky et al. showed that deep Convolutional Neural Networks could wiretap breakthrough performance on image classification tasks, which had many applications in other fields, including agriculture (Krizhevsky, Sutskever, & Hinton, 2012). This architectural innovation made it possible to automatically learn features, thus removing the description of features crafted manually.

The development of PlantVillage dataset was a significant breakthrough in AI studies in agriculture. Mohanty et al. gathered more than 54,000 images of 38 types of crops and 14 types of plant diseases, with a transfer learning accuracy of more than 99 percent with ImageNet-trained models. (Mohanty, Hughes, & Salathé, 2016). But behind these impressive results, were the fatal weaknesses. The photos of plantvillage have been taken in controlled laboratory conditions with standard backgrounds and standard lighting-conditions, which are not usually found in a real agricultural environment. The threat of domain shift was also noted in studies that tried to apply PlantVillage-trained models to field images, where accuracy dropped by over 30 per cent, indicating the domain shift issue (Maree, Fox, Al Rwahnih, Boonham, & Candresse, 2018).

investigated more advanced CNN designs to classify the diseases. He et al. proposed Residual Networks (ResNet), which used skip connections to make networks very deep, which has state-of-the-art performance on ImageNet (He, Zhang, Ren, & Sun, 2016). Proposed by Huang et al., DenseNet, which links all layers to each other in a feed-forward manner, encourages the use of features but minimizes the number of parameters (Huang, Liu, Van Der Maaten, & Weinberger, 2017). To reduce the computational cost, MobileNet presented depthwise separable convolutions to minimize the computational cost by an order of magnitude, making the use of depthwise separable convolution possible in real time on mobile devices (Howard, et al., 2017). EfficientNet systematically scaled network width, depth, and resolution, achieving superior accuracy-efficiency trade-offs (Tan & Le, 2019). They were broadly used to detect plant diseases, but the majority of the research still used PlantVillage, or a comparable controlled dataset.

Understanding these weaknesses, more realistic standards appeared. Introduced in the year 2020, the Plant Pathology 2020 FGVC7 challenge presented images of apple leaf captured in the field conditions with complex backdrops and varying lightings which are more reflective of the actual agricultural setting (Thapa R. , Zhang, Snavely, Belongie, & Khan, 2020). The 2021 FGVC8 challenge, based on this, enabled multi-label classification, which aligns with the biological fact that most plants can have several diseases simultaneously (Thapa R. , Zhang, Snavely, Belongie, & Khan, 2021). This development signifies maturation of the field towards its applicability.

Vision Transformers changed the basis of computer vision. Dosovitskiy et al. established that transformer-based architectures were capable of effectively adapting to image classification through viewing images as patch-sequence models and using self-attention mechanisms, and also competed with state-of-the-art CNNs on ImageNet (Dosovitskiy, et al., 2020). The quadratic nature of ViT and its large-scale pretraining made it difficult to use practically.

Liu et al. have introduced Swin Transformer which calculates self-attention in local windows and uses the shifting window schemes to facilitate cross-window interactions with a linear computational complexity and hierarchical feature representation (Liu, et al., 2021). In many vision tasks, classification, detection and segmentation Swin performed better. The hierarchical structure of it offers multi-scale characteristics useful in the process of identifying the symptoms of the disease at varying spatial levels. Although it has been widely used in the general computer vision field, the application of Swin to general computer vision agricultural disease detection questions has not been explored extensively. The revolution of transformers led to review of CNN design. Liu et al. presented ConvNeXt, which showed that CNNs might be as good as transformers with some modernization of the architecture (large 7x7 kernels and LayerNorm, GELU activations and inverted bottlenecks) (Liu, et al., 2022). This undermined accounts according to which the superiority of transformers was based essentially on the self-attention and indicated numerous benefits obtained by means of training processes and architectural considerations that could be used in either paradigm. ConvNeXt followed behind Swin Transformer with competitive accuracy in ImageNet whilst being convolutional efficient (Liu, et al., 2022). This challenged narratives that transformers' superiority stemmed fundamentally from self-attention, suggesting many gains derived from training procedures and architectural details applicable to both paradigms. ConvNeXt achieved competitive accuracy with Swin Transformer on ImageNet while maintaining convolutional efficiency (Liu, et al., 2022).

Multi-label classification, in which a sample is allowed to be classified into several classes at once, needs to take other methods than single-label classification. Independent application of Binary Cross-Entropy loss, with or without class-specific weights to deal with label imbalance, has become a standard practice (Lin, et al., 2017). To solve the problem of the imbalance in classes, Lin et al. introduced Focal Loss that down-weights well-classified examples and allows models to concentrate on the challenging ones (Lin, Goyal, Girshick, He, & Dollár, 2017). Multi-label methods have already been useful in medical imaging, where the patients have multiple conditions simultaneously, and these methods help to give the correct precedents concerning multi-disease plant situations. Nevertheless, single-label classification is still widely used in most plant disease systems.

The main objective of comparative analysis of CNN and transformer architectures has concerned general vision metrics. ImageNet-based systematic comparisons indicate that optimised CNNs such as ConvNeXt and transformers such as Swin get relatively little performance difference (Liu, et al., 2022). General-purpose benchmarks might not be indicative of agricultural imagery properties, where fine-grained textures, variation of colors and multi-scale appearances may be biased to one method. A direct comparison of ConvNeXt and Swin on agricultural multi-label tasks is still not present in the literature.

2.3 Research Gap

Although much has been achieved in computer vision in the field of agriculture, there are still various gaps that are evident in the literature. To begin with, most of the studies still test models on artificial, laboratory-controlled data such as on the PlantVillage, with even backgrounds and artificial lighting, which does not transfer to the field (Mohanty, Hughes, & Salathé, 2016) (Barbedo, 2018). The challenges of 2020 and 2021 in the field of Plant Pathology showed realistic data (Thapa R. , Zhang, Snavely, Belongie, & Khan, 2020) (Thapa R. , Zhang, Snavely, Belongie, & Khan, 2021), there has been little comprehensive architecture comparison on these challenges. The majority of the recent literature still dwells on a single architecture or comparison limited to the traditional CNNs without regard to the transition to the modern Transformer-based backbones.

Second, the single-label classification does not consider the biological fact of co-occurring plant diseases. In agro-ecosystems of nature, several pathogens often invade the same tissue, especially where there are environmental (Savary, et al., 2019). Single-label methods compel models to only classify the most dominant condition and ignore the precious diagnostic data. Although the Plant Pathology 2021 dataset has multi-label annotations(Thapa R. , Zhang, Snavely, Belongie, & Khan, 2021), the use of advance multi-label loss functions and thresholding techniques are not studied as much as the conventional softmax classification.

Third, current comparisons of CNNs with Transformers are mostly focused on general vision tasks such as ImageNet where both models can do similarly well on the high-level task (Dosovitskiy, et al., 2020). However, the domain specific features of plant disease imagery, including coarse grain morphology of symptoms textures, variations in colour that are not easily seen, and multi-scale features, can be biased towards one architectural philosophy but not the other. The significance of local feature extraction (also common to CNNs) versus global context modeling (also common to Transformers) has not been empirically validated in any complex agricultural setting.

Fourth, most architectural comparisons focus on prediction performance and ignore computational performance, which is essential to implement them on resource-limited agricultural edge devices (Saleem, Potgieter, & Arif, 2019). It is important to practitioners to understand the trade-off between accuracy, inference speed and the size of parameters but extensive assessments including such measures are uncommon. Lastly, recent developments of ConvNeXt and Swin Transformers as exemplary convergent design philosophies have not been pursued in precision agriculture. The given thesis fills these gaps directly by means of the systematic comparison of the given modern architectures on the multi-label apple disease classification.

2.4 Summary

Deep learning has since progressed in the form of early CNNs on controlled datasets to current designs such as Swin and ConvNeXt that aim to simulate the realities of fields. Nevertheless, there are still serious gaps on multi-label classification, computational efficiency, and direct architectural comparisons. This thesis goes about these drawbacks by critically assessing these contemporary models using field-based data.

CHAPTER 3 METHODOLOGY

3.1 Introduction

The study takes a systematic channel of data collection to the process of model testing as shown in Figure 3.1. The dataset that contains real-world images of apple leaves is converted into a 6-class multi-label framework, which, in turn, allows detecting the presence of diseases in co-occurrence (Thapa R. , Zhang, Snavely, Belongie, & Khan, 2021). The dataset is stratified into 80:10:10 to create balanced distribution between training set, validation and test set.

ConvNeXt-Tiny and Swin-Tiny are both initialized using ImageNet-22k pretrained weights and both optimized with the same settings. Large-scale datasets allow transfer learning to promote successful adaptation to specialized agricultural tasks with limited labels (Müller, Jaeger, & Hanewinkel, 2019) The training employs AdamW with cosine learning rate scheduling and as well as early stopping relying on the validation macro F1-score. The augmentation of data is done on training data only.

The comprehensive multi-label metrics applied in model evaluation on the test set are macro F1-score (primary), precision, recall, and exact match ratio. Inference speed, model size, and gpu memory requirement determine the level of computational efficiency and can be used to offer practical deployment advice. Grad-CAM visualization confirms that models pay attention to biologically significant areas of disease and shows a difference in attention patterns in architectures.

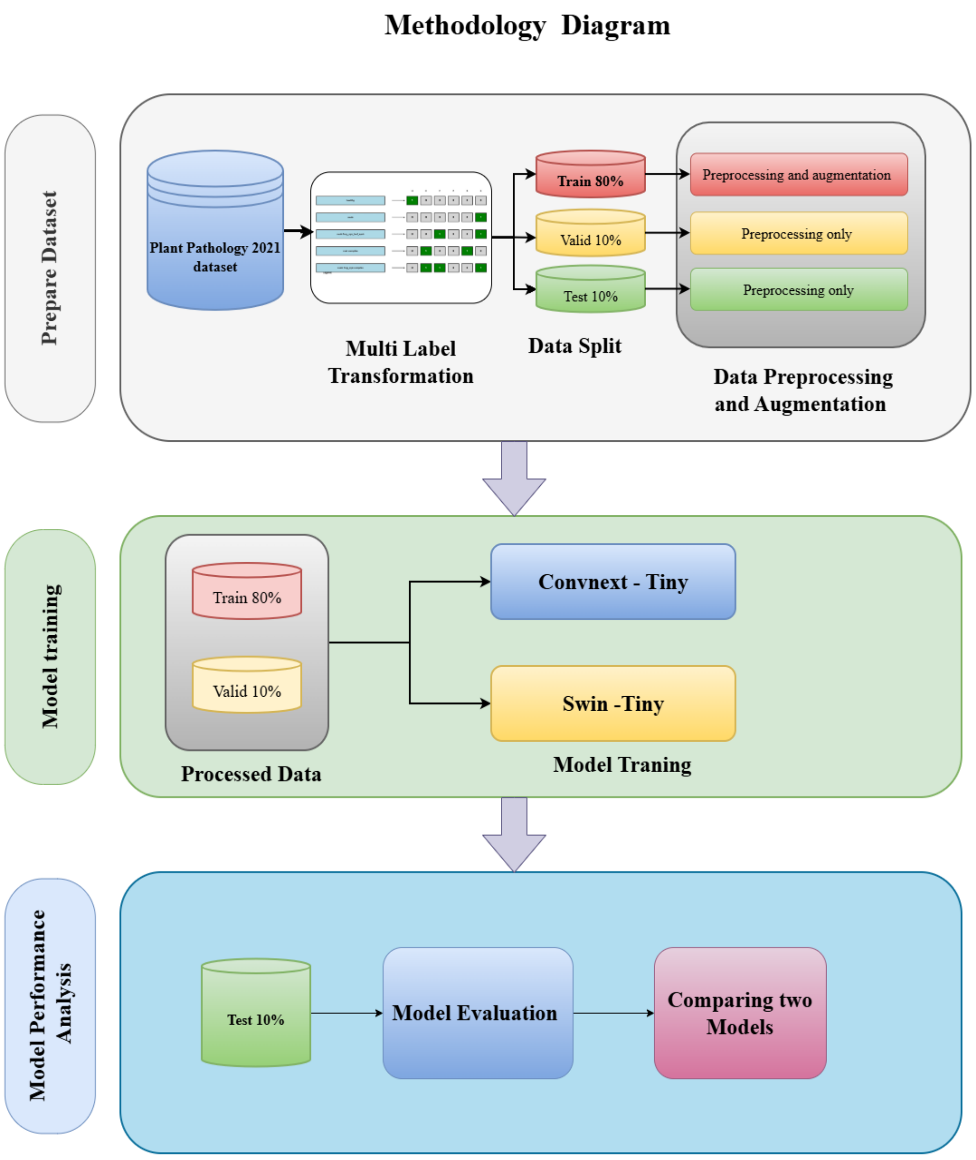


Figure 3.1: Methodology Diagram

3.2 Dataset Description

The Plant Pathology 2021 FGVC8 data has more than 20,000 RGB images of apple leaves taken in the field, which were annotated in the Fine-Grained Visual Categorization workshop (Thapa R. , Zhang, Snavely, Belongie, & Khan, 2021). Unlike laboratory databases like PlantVillage, these images reflect the complexities of real world agriculture like complex backgrounds, changing lighting conditions, varying camera angles, different stages of leaf maturity and partial occlusion. This type of realism makes the data set especially appropriate in the evaluation of deployment-ready models.

The initial annotation scheme had 12 disease categories, of which six (healthy, scab, frog eye leaf spot, rust, powdery mildew and complex) were single-disease classes and six (mixture of 2 or more diseases) were multi-disease combinations. All 12 classes are represented by samples in figure 3.2. Figure 3.3 shows the full distribution of the 12 original classes whereas Figure 3.4 shows the patterns of co-occurrences that provides a visualization of disease combinations in the dataset. This multi-label structure is biologically relevant where various pathogens may be capable of infecting plant tissues at a time, although it is also suitable in assessing multi-label classification methods.

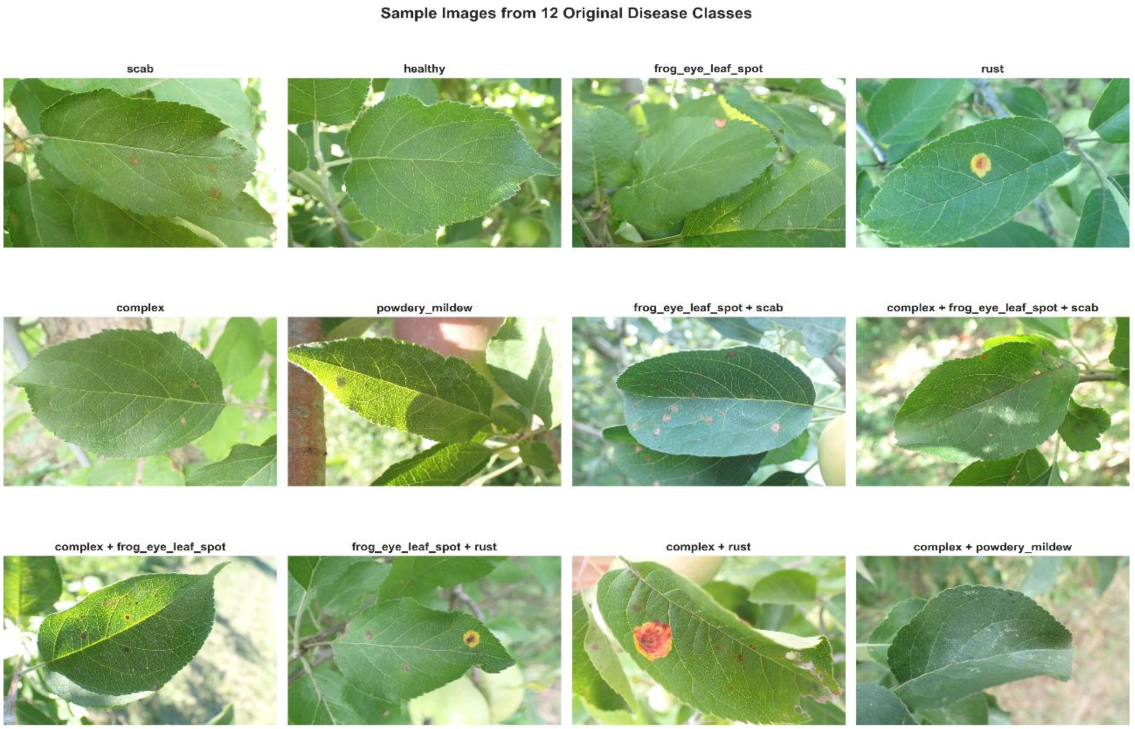


Figure 3.2 : Sample images from all 12 original classes

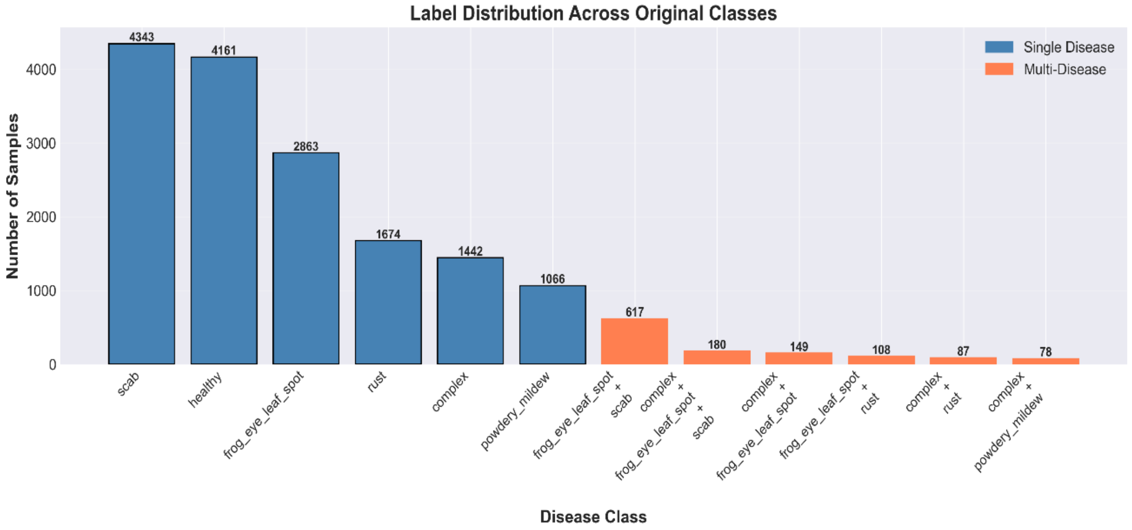


Figure 3.3 : Label Distribution across original classes

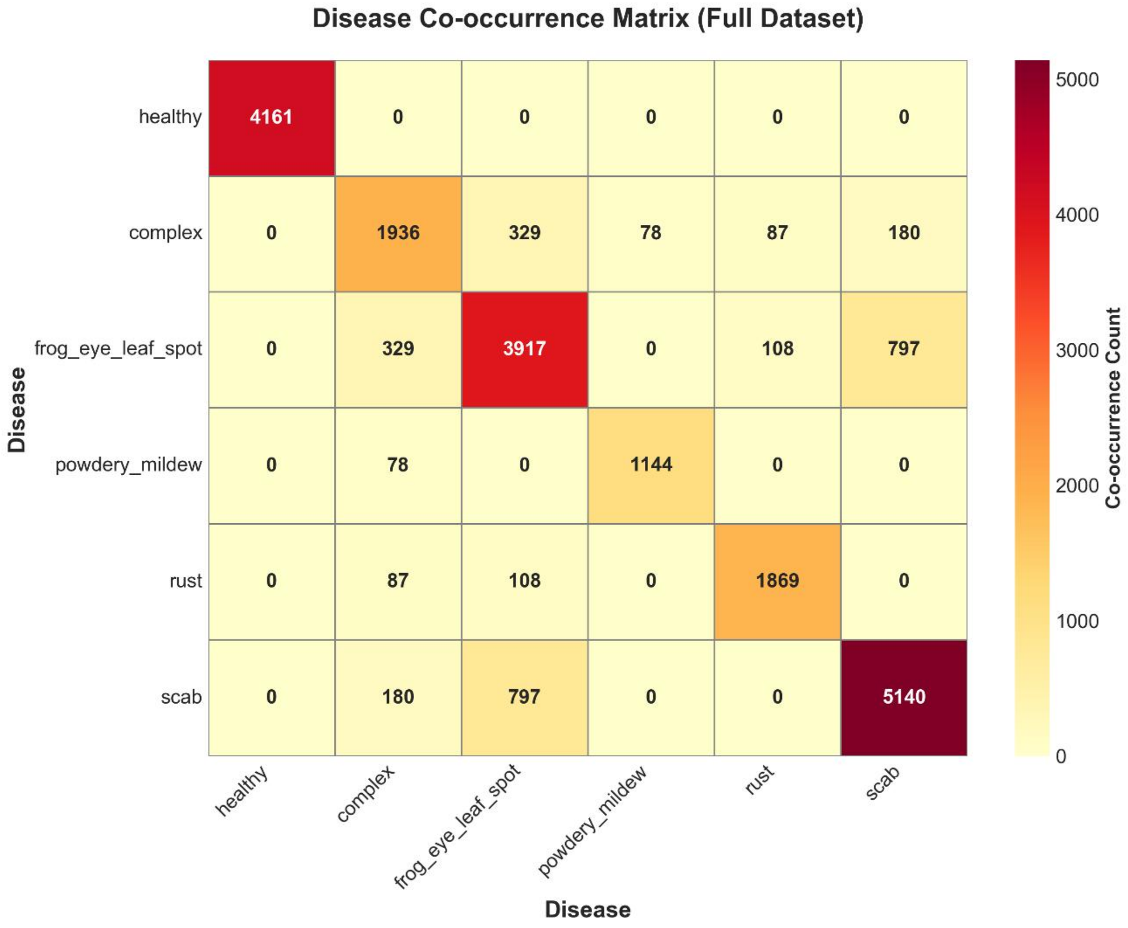


Figure 3.4 : Co-occurrence matrix

3.3 Data Preprocessing and Preparation

3.3.1 Multi-Label Transformation

The The initial 12-class system was transformed into a multi-label system with six independent binary labels, which included healthy, complex, frogeyeleaf spot, powdery mildew, rust and scab. This transformation considers each disease as a separate binary classification problem, and allows simultaneous prediction of many co-occurring diseases.

Multi-hot encoding The images are represented by means of a binary vector, which has a length of six, with a 1 indicating the presence of disease and 0, the absence. An example is that the word scab is coded as [0, 0, 0, 0, 0, 0, 1] and scab frogeyeleaf spot complex is coded as [0, 1, 1, 0, 0, 1]. The mutually exclusive element between the classes of health and disease is that in case of healthy = 1, the rest of the positions will be zero. However, co-occurring diseases can be in any combination where samples can be having one up to three active labels.

This transformation with an example of mapping of original label to multi-hot vectors is shown in Figure 3.5. In contrast to multi-class classification, this versatile method of classification determines a combination of diseases to facilitate sound agriculture decision-making.

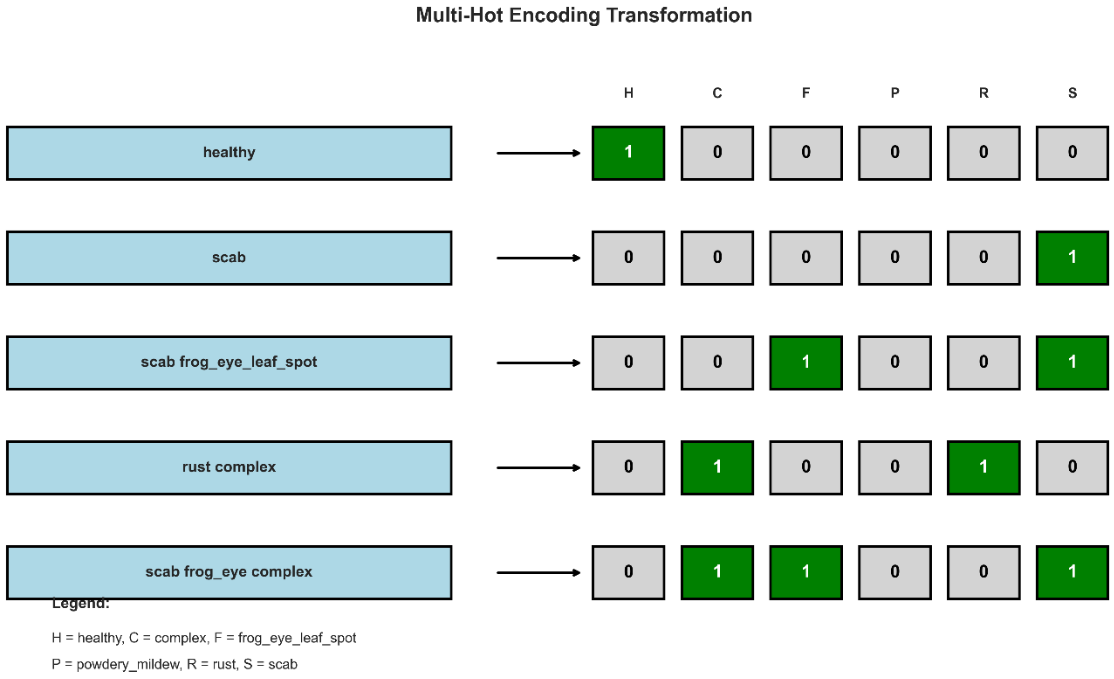


Figure 3.5 : Multi-hot encoding transformation

The multi-hot encoding transformation provide six different classes; the data distribution across these classes is illustrated in Figure 3.6.

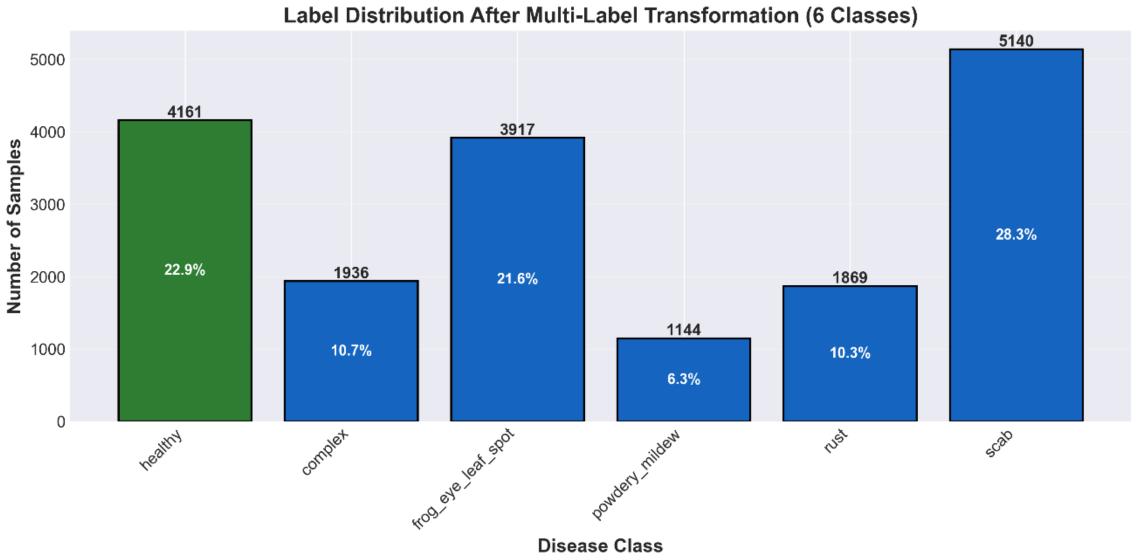


Figure 3.6 : Label distribution after multi-label transformation

3.3.2 Dataset Splitting and Class Balancing

The Stratified random sampling with 80 10 10 ratio was used to divide the data set into 13440 training, 1680 validation and 1680 test images. The stratification was done based on combinations of diseases as opposed to a single disease and thus it preserved the distribution of co-occurring patterns across splits as well as co-occurring patterns are present in all sets which are rare. The imbalance in the classes was alleviated by assigning classes positive weighted loss. Calculate the weights as :

$$positive\ weight = \frac{total\ samples - positive\ samples}{positive\ samples}$$

Table 3.1 reveals the counts of samples and weights. Powdery mildew was the most weighted (13.66×) because it is the least common disease but scab weighs the least (2.26×) because it is the most common disease. These were added to Binary Cross-Entropy with Logits loss whereby the model distributes enough learning capacity to infrequent diseases even though they have fewer samples. The split partitions distribution is plotted in Figure 3.7.

Table 3.1: Class distribution and positive weights

Disease	Samples	Percentage	Pos_Weight
scab	5,140	30.6%	2.26
healthy	4,200	25.0%	3.03
frog_eye_leaf_spot	3,900	23.2%	4.10
rust	1,900	11.3%	10.50
complex	1,800	10.7%	9.80
powdery_mildew	1,144	6.8%	13.66

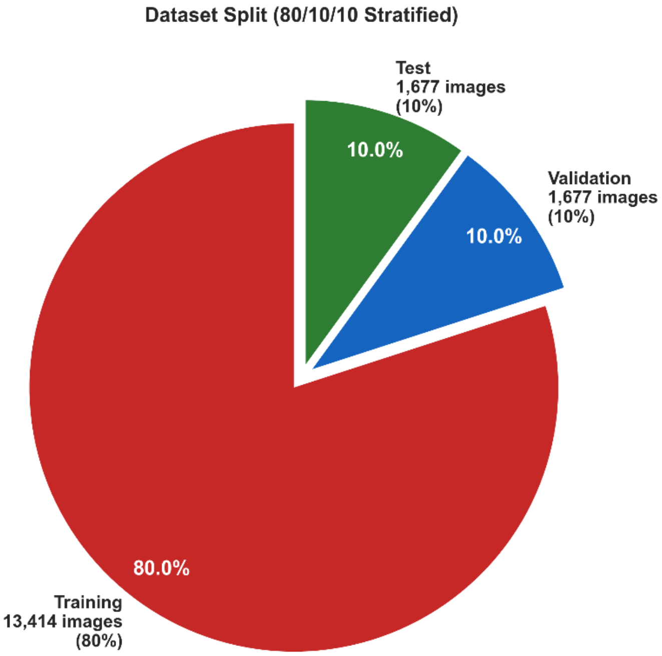


Figure 3.7: Data split visualization

3.3.3 Image Preprocessing and Augmentation

All images were resized to 224×224 pixels so that it can match with the model actual input. The ImageNet normalization process was applied to all RGB channels for pixel value normalization which made the data compatible with models that used pre-trained weights.

The training data received data augmentation through these five techniques which included Random Resized and Random Horizontal Flip for orientation invariance and

Random Vertical Flip for natural leaf positions and Random Rotation for camera angle simulation and Color Jitter for lighting and sensor variation simulation.

The validation and the test sets received only resizing and normalization, didn’t apply augmentation here so that it can ensure fair evaluation on unmodified data. The asymmetric strategy enables the model to learn features from different training environments which results in better performance during testing scenarios.

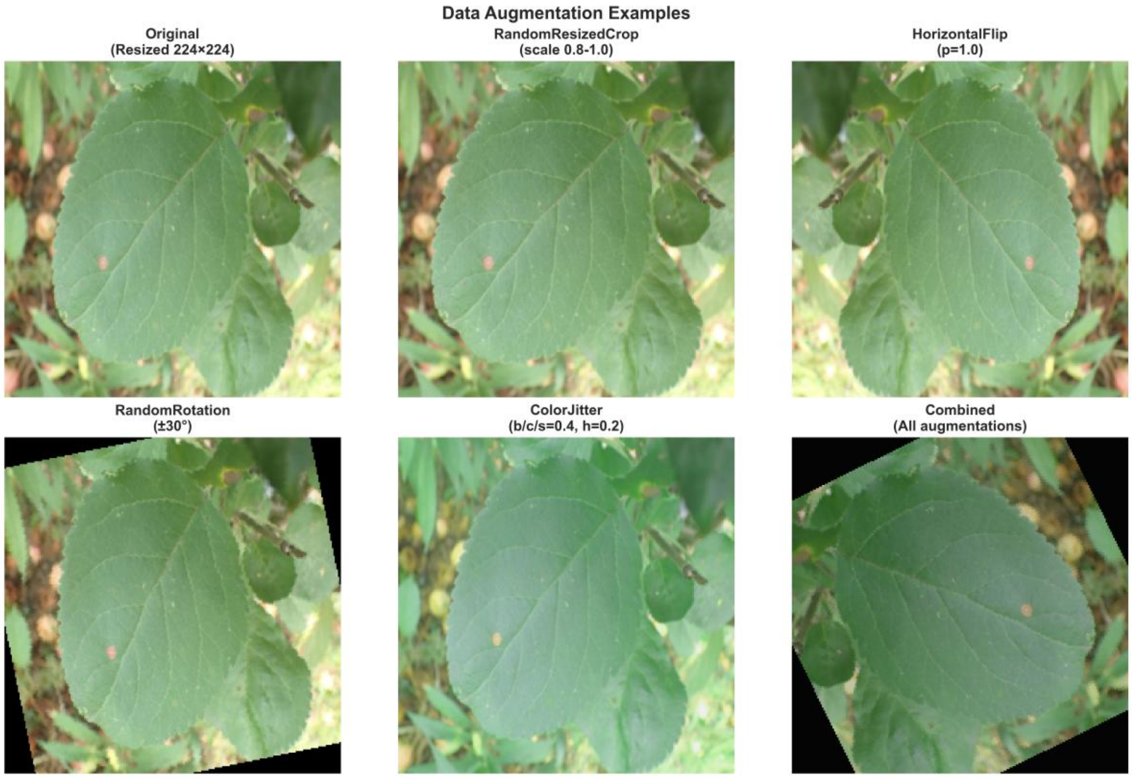


Figure 3.8: Data augmentation examples

3.4 Model Architectures

3.4.1 Model Selection Rationale

Two state of the art architecture have been chosen to be compared and contrasted, namely ConvNeXt -Tiny and Swin -Tiny, representing two different design paradigms in modern computer vision. ConvNeXt is an illustration of modernised convolutional neural networks, which apply principles of transformer-based design but inductive biases of convolutional bias are still present. The Swin Transformer represents a hierarchical vision transformer, which use shifted window attention to provide efficient global modelling. The two models share similarities in terms of their characteristics, which enable controlled comparison: both are roughly 28m parameters, both models are around 105-107MB in size and the input resolutions of both models are identical (224x 224 pixels). They were firstly pretrained on ImageNet-22k then fine-tuned on ImageNet-1k therefore offering powerful general feature representations to the transfer learning (Liu, et al., 2022). The smaller versions were selected based on performance and computational efficiency to make them appropriate to any real-world situation of agricultural implementation, including edge devices and mobile platforms. The models were ran with the help of the library TIMM (PyTorch Image Models), which provides standardised interfaces and pretrained weights in the various architectures (Wightman, 2019), thus, being reproducible and allowing a fair comparison under the same training.

3.4.2 ConvNeXt-Tiny Architecture

ConvNeXt-Tiny is a modern CNN model that learns to use the design of vision transformers but retains the convolutional operations (Liu, et al., 2022). The main design elements are inverted bottleneck blocks and 4x channel expansion, the replacement of the Batch Normalisation with Layer Normalisation to increase the stability of the training process, and GELU activation functions. The network has four stages that are hierarchical with a continually smaller spatial resolution (56, 28, 14, 7), and a channel dimension that is correspondingly larger (96, 192, 384, 768). The model was launched with ImageNet-

22k pretrained weights and ImageNet-1k fine tuned. Its configuration has a drop-rate of 0.2 to dropout regularisation of fully connected layers and a drop rate of 0.1 to use stochastic depth during training. Instead of the original 1000-class ImageNet classifier head, a linear layer that produces six-label logits was used, and the sigmoid activation was used as part of inference. The full model has a size of about 107MB in FP32 precision and it consists of 28 8 million parameters. Figure 3.9 is the ConvNeXt architecture, shows the specifications of the stages and block configurations.

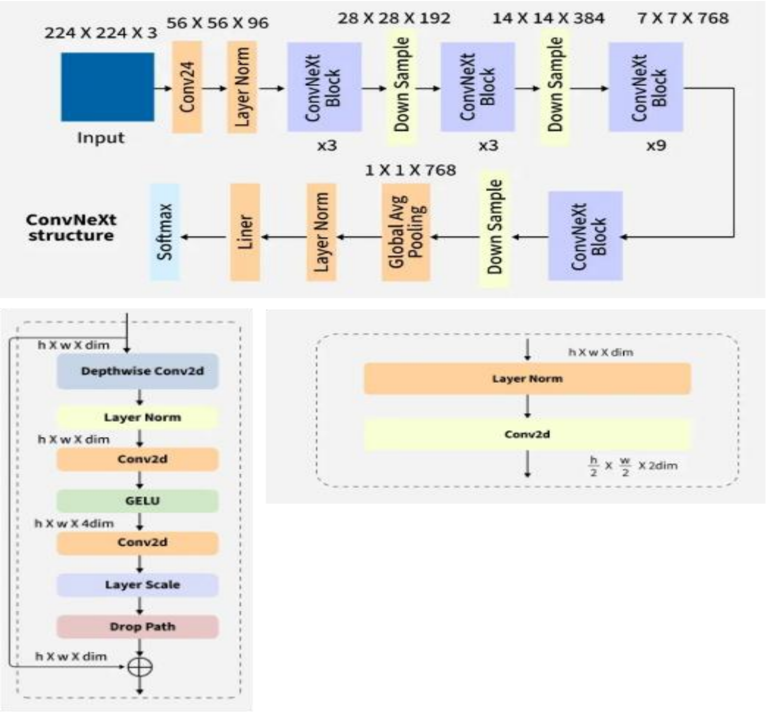


Figure 3.9: ConvNeXt-Tiny architecture diagram (Liu, et al., 2022).

3.4.3 Swin-Tiny Architecture

The Swin Transformer uses a hierarchic design, which involves self-attention in the local, non-overlapping windows, to reach linear computational complexity to build the multi-scale feature representations (Liu, et al., 2021). The architecture then starts with a 4 x 4 patch embedding. The partitions by shifted window mechanism partition maps into 7 7 windows to perform local self-attention which are then moved by three pixels to allow cross-window connection without compromising on efficiency. Patch-merging layers reduce the feature maps by merging the patches around them and enforcing linear projections in order to build hierarchical representations at four levels. The model was initialised with ImageNet -22k pretrained weights and fine-tuned on ImageNet -1k. The settings are a droprate of 0.2 and a drop-path rate of 0.2. The multi-head self attend uses at four stages. A linear head that outputs six classes multi-label was used instead of the original classifier head. Figure 3.10 illustrate the Swin Transformer architecture.

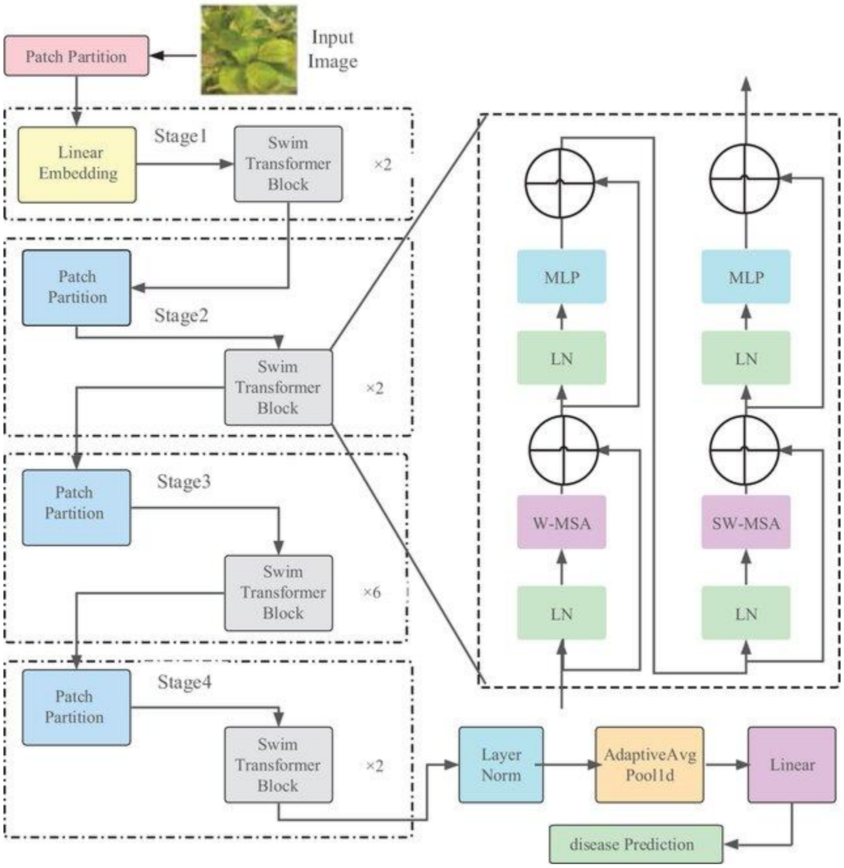


Figure 3.10: Swin-Tiny architecture diagram (Liu, et al., 2021).

3.4.4 Architecture Comparison

Table 3.2 provides the side-by-side comparison of ConvNeXt-Tiny and Swin-Tiny specifications. The basic difference is in their method of extracting features as ConvNeXt uses convolutional inductive biases based on explicit hierarchical down-sampling and Swin uses global self-attention in local windows with learned positional biases. The larger kernels of ConvNeXt (7 x 7) partly close the gap in the receptive fields between transformers with that of Swin, which results in cross-region communication using shifted windows. The two architectures produce hierarchical multi-scale characteristics that are appropriate to dense prediction. The similar number of parameters and model sizes used by them make them computationally equivalent.

Table 3.2: Architecture specifications comparison

Specification	ConvNeXt-Tiny	Swin-Tiny
Architecture Type	Modernized CNN	Hierarchical Transformer
TIMM Identifier	convnext_tiny.fb_in22k_ft_in1k	swin_tiny_patch4_window7_224.ms_in22k_ft_in1k
Parameters	~28M	~28M
Model Size (FP32)	107MB	105MB
Input Resolution	224×224	224×224
Patch/Kernel Size	7×7 depthwise	4×4 patch embedding
Stages	4 (56→28→14→7)	4 (56→28→14→7)
Attention Type	None (pure conv)	Shifted window (7×7)
Normalization	LayerNorm	LayerNorm
Activation	GELU	GELU
Drop Rate	0.2	0.2

3.5 Training Configuration

3.5.1 Optimization and Loss Function

The objective of the model is to optimize the following function:

The size of a batch used was 24, which was optimised to the memory that was available on the GPUs. AdamW optimiser was used where the learning rate was 5 x 10 -5, a weight decay of 0.05, and two beta parameters (0.9, 0.999) and epsilon 1 - = (Loshchilov, 2019; Hutter, 2019). AdamW separates the idea of weight decay and gradient updates, and thus it offers better regularisation compared to conventional Adam. Learning -rate scheduling was applied using CosineAnnealingLR with a minimum learning rate of 1 -10-6, which smoothly decreased the learning rate over the course of the training. The learning rate was gradually raised in the initial three epochs in a linear manner starting at zero and stabilised the early training dynamics. The loss used BCEWithLogitsLoss (Binary Cross Entropy with logits), as it includes class-specific positive weights taken in Table 3.1 to deal with gross class imbalance. This loss combines both sigmoid activation and binary cross-entropy in a numerically stable loss and uses independent binary classification on each of the six disease labels. Gradient clipping to a maximum norm of 1.0 to prevent exploding gradients, dropout (rate=0.2) and stochastic depth through drop-path were all regularisation methods. Training mixed-precisions was disabled because stability was a consideration in the windows platform.

3.5.2 Training Procedure

The The warm-up phase (epochs 1-3) was a linear learning-rate warm-up, where the rate was raised to 0 correspondingly and in a constant proportion to 5X10-5. The cosine-annaling learning-rate decay was used in the main training (epochs 4-50). Evaluation of the model performance on the validation set was done after every epoch, and all the multi-label measurements were calculated. The checkpoint with the maximum validation macro F1 -score was stored as the optimal model to be used in the future during test analysis. Checkpoints were periodically saved after every five epochs to make recovery during training possible, in case of need.

Table 3.3: Training hyperparameters summary

Hyperparameter	**ConvNeXt-Tiny**	**Swin-Tiny**
Batch Size	24	24
Learning Rate	5×10^{-5}	5×10^{-5}
Weight Decay	0.05	0.05
Optimizer	AdamW	AdamW
LR Scheduler	CosineAnnealingLR	CosineAnnealingLR
Min LR	1×10^{-6}	1×10^{-6}
Warmup Epochs	3	3
Max Epochs	50	50
Early Stopping	10 patience	10 patience
Drop Rate	0.2	0.2
Drop Path Rate	0.1	0.2

3.6 Evaluation Metrics

3.6.1 Performance Metrics

Detailed multi-label classification measures were used as evaluations of model performance. The general measures were Macro F1 -score (major measure), Micro F1 -score, Weighted F1 -score, label -wise accuracy, exact match loss, and Hamming loss (Zhang & Zhou, 2014). MacroF1-score F1-scores are also averaged over all the six diseases, and thus the rare diseases are considered as important as common diseases, which is necessary in the presence of unbalanced data. It was selected as the major comparison measure since in agricultural practices there is a need to have a sound checking of all diseases irrespective of the prevalence. Micro F1-score summarizes true positives, false positives, and false negatives of all labels and then calculates F 1, which focuses on the majority classes. F1 -scores averages per-class F1 -scores weighted by class support (number of positive samples). Label-wise accuracy is the percentage of correctly called labels in all samples and classes. Exact match ratio (subset accuracy) is the harshest measure, where all six labels are correctly predicted at once - the proportion of samples on which they are. The Hamming loss is used to determine the mean rate of mislabeling samples. Per disease measures were precision (positive predictive value), recall (sensitivity / true positive rate), F1-score (harmonic mean of the two) and specificity (true negative rate) of each of the six diseases. The granular metrics allow us to see which diseases each architecture identifies best thus what their possible complementary strengths would be.

3.6.2 Computational Metrics and Evaluation Protocol

Four key metrics of computational efficiency were measured to determine the deployability of the system under both cloud resources that are resource-rich and edge devices with very limited resources. The average time to process one batch (24 images) was recorded and used to calculate per-image latency and throughput (images per second), inference time. Also, model size (in megabytes, FP32 precision) was noted to reveal how much storage is needed and how much memory-bandwidth is used during loading. These efficiency metrics combined with predictive performance together may be used to select architectures based on some applicational constraints. The assessment plan was the means of rigorous and objective evaluation. The most effective checkpoint of the model (selected by the largest validation macro F1 -score) was loaded and configured to be in the evaluation mode where stochastic regularization methods like dropout and droppath were disabled. The predictions were obtained in the held-out test set without the computation of the gradients to reduce memory consumption. The raw model logits were smoothed using a sigmoid activation and probabilities binarized with a set threshold of 0.5. These binarized performance measures were calculated using these results compared to the ground-truth labels. Stratified data splitting and disjoint training, validation and test sets were used to ensure that there was no leakage of data hence good performance estimates were obtained.

3.6.3 Interpretability Analysis with Grad-CAM

Interpretability analysis based on Gradient-weighted Class Activation Mapping (Grad-CAM) was performed to ensure that the models were based on biological relatable symptoms of the disease, and not based on artifacts or noise in the background(Selvaraju, et al., 2017). Grad-CAM produces visual explanations by calculating the score differences of the target class with the respect to the feature maps of the final convolutional layer, resulting in a rough localisation map that identifies the most discriminative regions to make the prediction. Architectures were somewhat different in the way they were implemented. In the case of ConvNeXt-Tiny, the activations were directly obtained in the last normalisation layer (normpre) before the classification head. In the case of Swin-Tiny, activations were taken out of the last Layer- Norm.

3.7 Summary

This chapter described how ConvNeXt-Tiny and Swin-Tiny can be compared with the help of Plant Pathology 2021 dataset modified into a multi-label six-class system. In order to counteract data imbalance, stratified splitting, class-weighted loss, and robust augmentation were used in the study. Both models (~28 million parameters) used the same training settings such as AdamW optimisation and cosine scheduling, which provided fair play of the experiment. Multi-label performance measurements, computational efficiency measurements and Grad-CAM interpretability analysis were evaluated. Through these approaches, the strict foundation is laid down to the results provided in Chapter 4.

CHAPTER 4 RESULTS AND DISCUSSION

4.1 Introduction

It is a comprehensive experimental analysis of two different architectures, ConvNeXt - Tiny and Swin -Tiny, on the demanding challenge of multi-label plant disease classification. Each of the two architectures was then put under a standardized experimental protocol using the same training settings and hyperparameter choices to provide a strict and objective comparison. A test with 1 677 held out test images was used to test performance and determine how well the results could be generalised to unseen data. Convergence stability in training, quantitative measures of classification, and visual interpretability analysis are the dimensions of analysis analyzed in the results published in this paper. The chapter is closed with a critical comparative analysis which summarises these results and clarifies the comparative merits of each of the models in the diagnosis of agricultural diseases.

4.2 Training Results

4.2.1 ConvNeXt-Tiny Training

After 5.22 h of training, ConvNeXt - Tiny reached its highest level of performance at the 35 th epoch with a validation macro- F1 of 0.9220. The loss curves of training and validation are shown in figure 4.1. The training loss decreased to 0.31 to 0.08, and the validation loss was almost 0.16. Ten consecutive epochs without any improvement led to the premature termination of training.

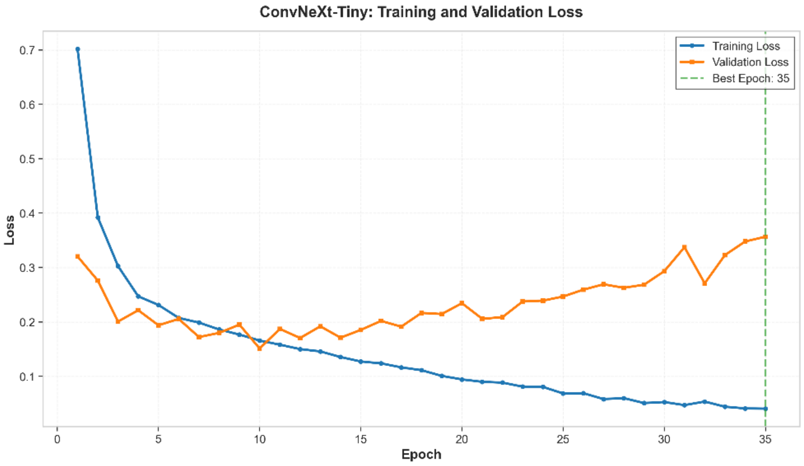


Figure 4.1: ConvNeXt-Tiny training and validation loss curves

4.2.2 Swin-Tiny Training

Swin-Tiny achieved its best performance at epoch 33, with a validation macro-F1 score of 0.9253 after 6.66 hours of training. Figure 4.2 shows the paths of the losses. There were similar convergence patterns to ConvNeXt, but the final validation loss was slightly lower (0.1589 vs. 0.1653). Figure 4.9 shows a direct comparison of training between architectures.

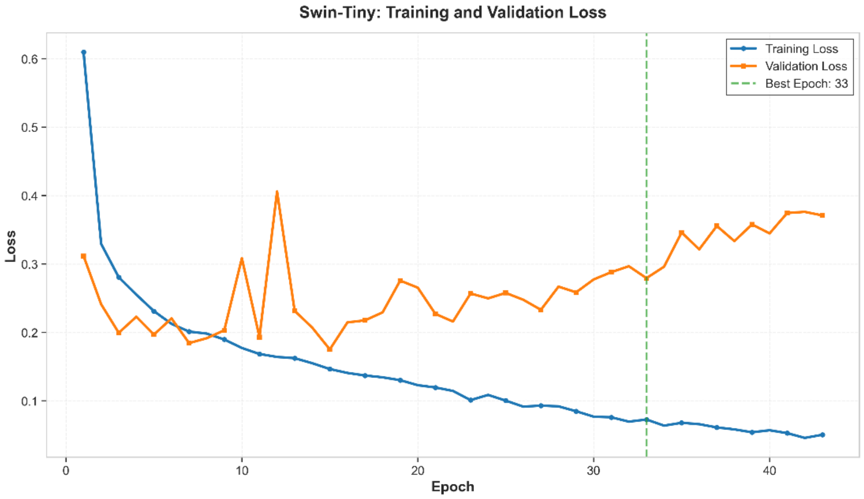


Figure 4.2: Swin-Tiny training and validation loss curves.

Figures 4.3 and 4.4 illustrate the validation accuracy of the ConvNeXt-Tiny and Swin-Tiny models, respectively.

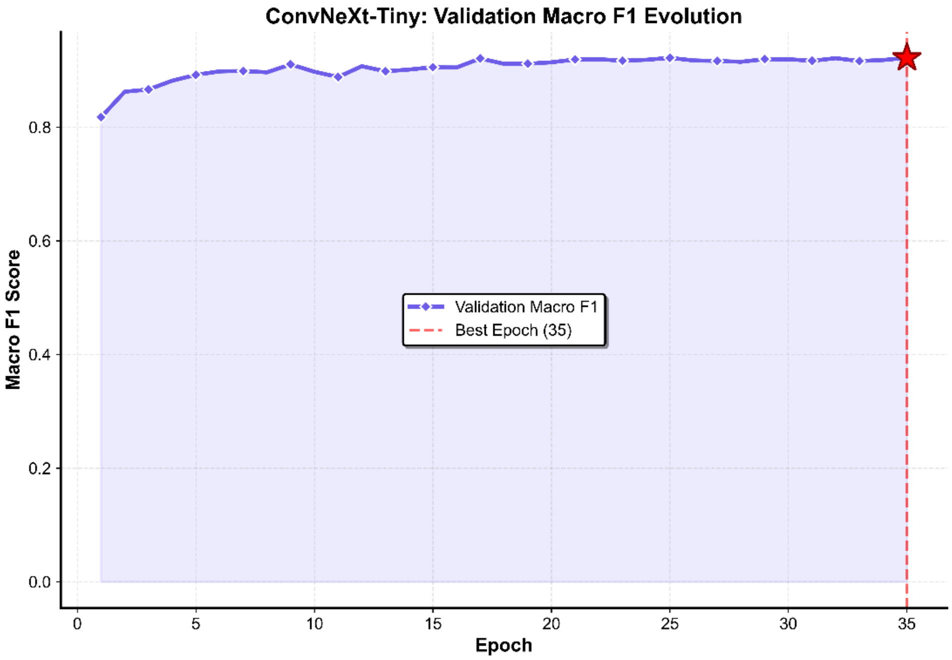


Figure 4.3: ConvNeXt-Tiny Validation macro-F1

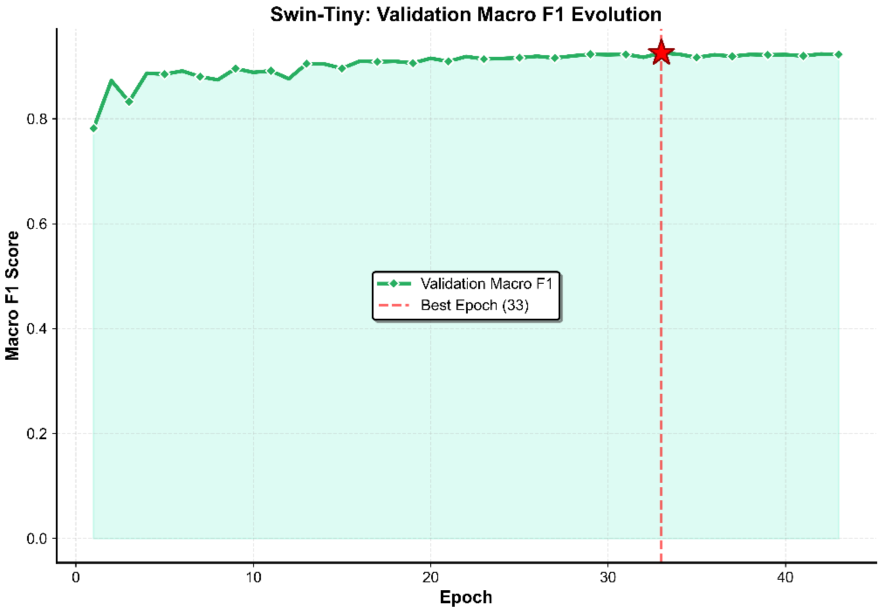


Figure 4.4: Swin-Tiny Validation macro-F1

4.3 Test Set Performance

4.3.1 Overall Metrics

The results for ConvNeXt-Tiny are shown in Table 4.1: a macro F1 of 0.9058, an accuracy of 96.90%, and an exact match ratio of 88.61%. Table 4.2 shows the results for Swin-Tiny: a macro F1 score of 0.9066, an accuracy of 96.79%, and an exact match ratio of 87.84%. Both models show that they can do multi-label classification well and compete with each other.

Table 4.1: ConvNeXt-Tiny test performance.

Metric Category	Specific Metric	Value
F1 Scores	Macro F1	0.9058
	Micro F1	0.9150
	Weighted F1	0.9150
Precision	Macro Precision	0.9001
	Micro Precision	0.9066
Recall	Macro Recall	0.9121
	Micro Recall	0.9235
Overall Performance	Accuracy	96.90%
	Exact Match Ratio	88.61%
	Hamming Loss	0.0310

The Swin-Tiny model achieved its best results on multiple critical evaluation metrics which were tested during the assessment process. The model produces 96.79% accurate results when it runs its forecasting system. The Recall scores (0.9234 for Macro and 0.9340 for Micro) show better performance than the Precision scores which reach approximately 0.89. This means that the model is very good at finding relevant instances, but it might give a few false positives. The Exact Match Ratio of 87.84% indicates that the model correctly predicts all labels for most test samples.

Table 4.2: Swin-Tiny test performance.

Metric Category	Specific Metric	Value
F1 Scores	Macro F1	0.9066
	Micro F1	0.9131
	Weighted F1	0.9136
Precision	Macro Precision	0.8915
	Micro Precision	0.8932
Recall	Macro Recall	0.9234
	Micro Recall	0.9340
Overall Performance	Accuracy	96.79%
	Exact Match Ratio	87.84%
	Hamming Loss	0.0321

4.3.2 Per-Disease Performance

Precision, recall, and F1-scores for each disease are shown in Figures 4.5 and 4.6. Healthy and powdery_mildew are the labels that perform best, followed by rust, scab, and frog_eye_leaf_spot. For both architectures, the complex multi-disease class exhibits the lowest performance.

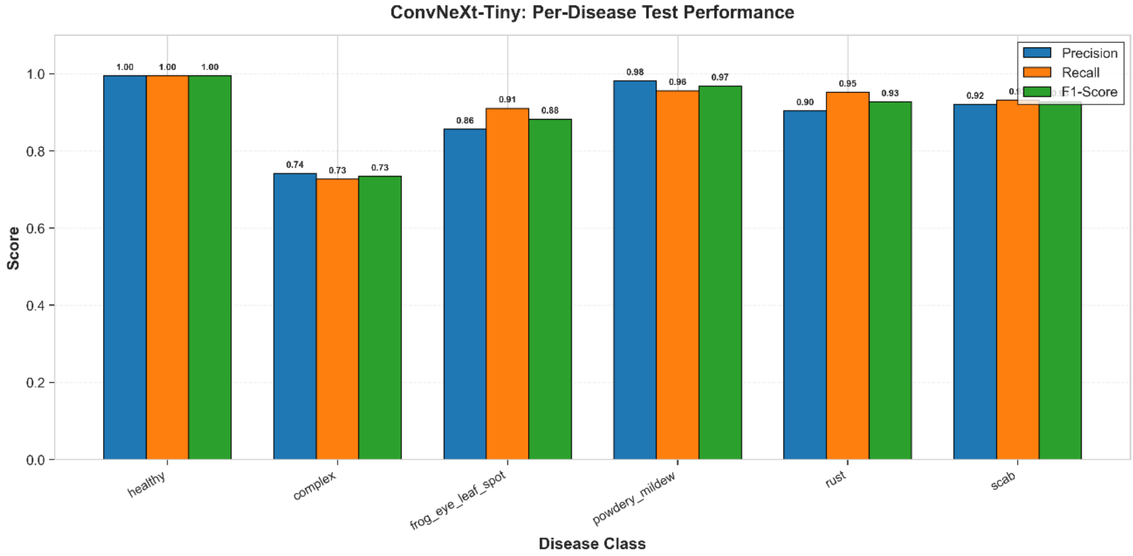


Figure 4.5: ConvNeXt-Tiny per-disease performance

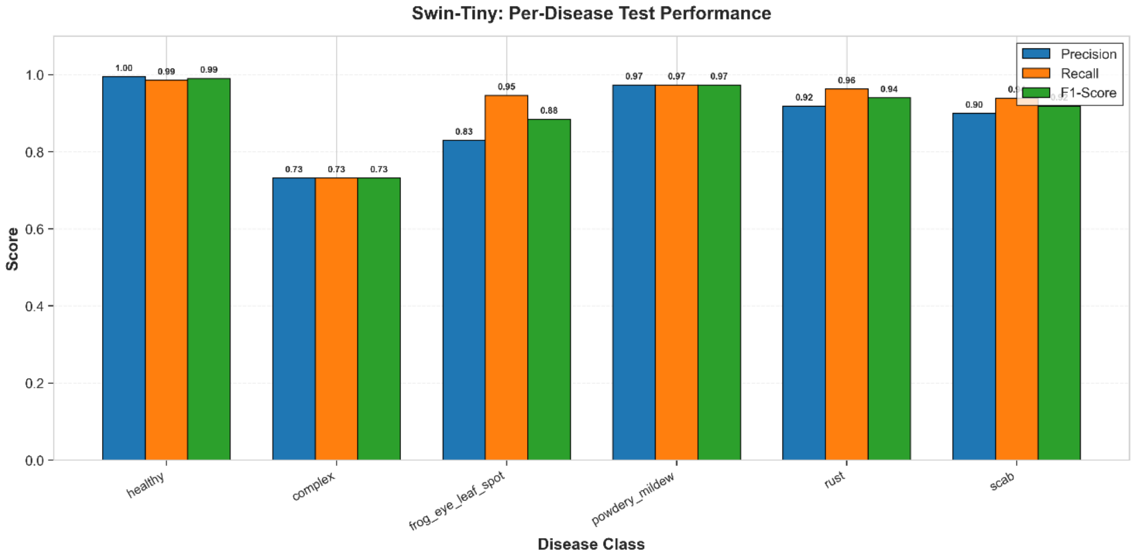


Figure 4.6: Swin-Tiny per-disease performance

For every disease, confusion matrices are shown in Figures 4.7 and 4.8. Near-perfect discrimination is achieved by healthy classification, while the highest level of confusion is demonstrated by complex classification.

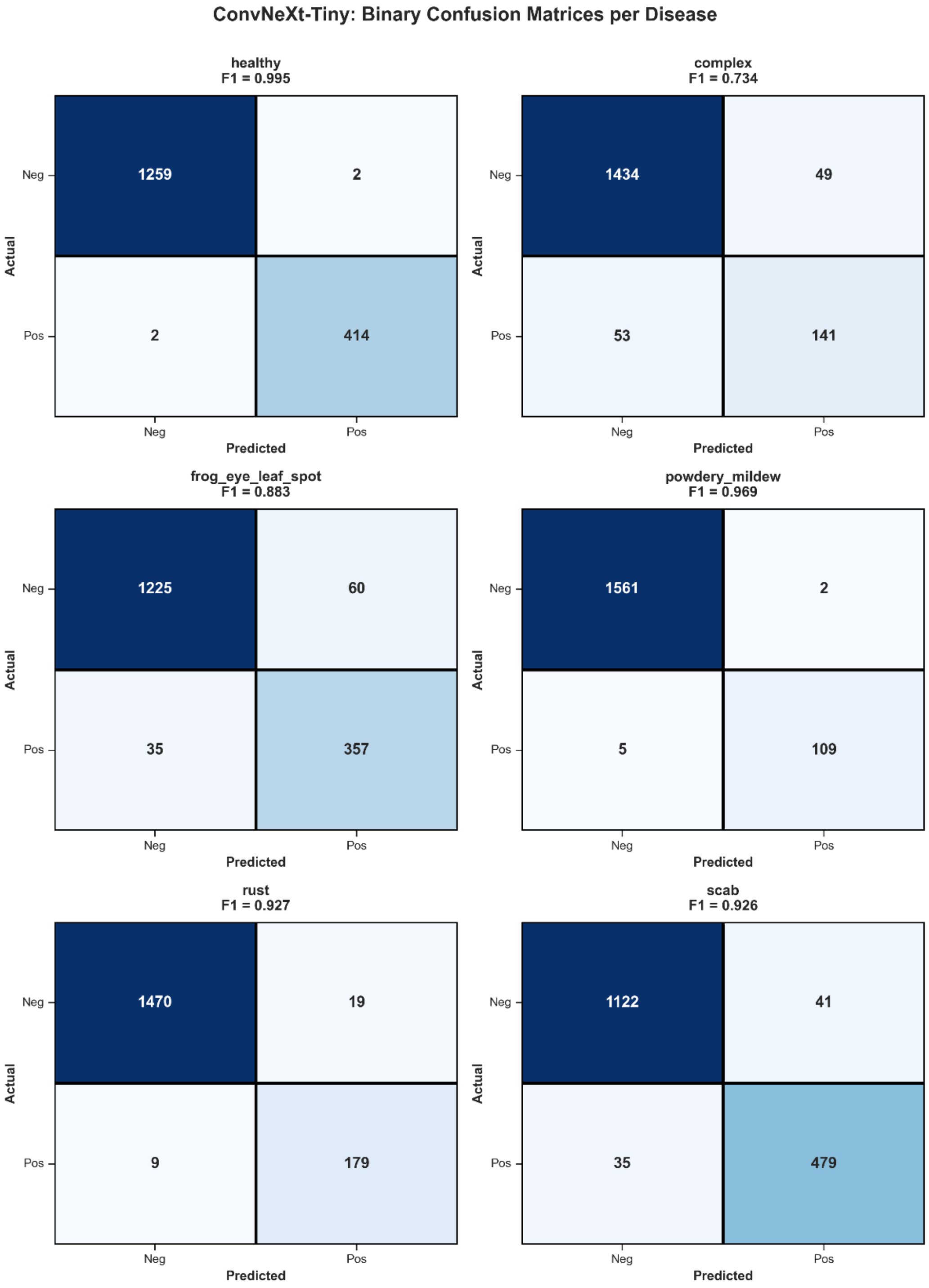


Figure 4.7: ConvNeXt-Tiny confusion matrices

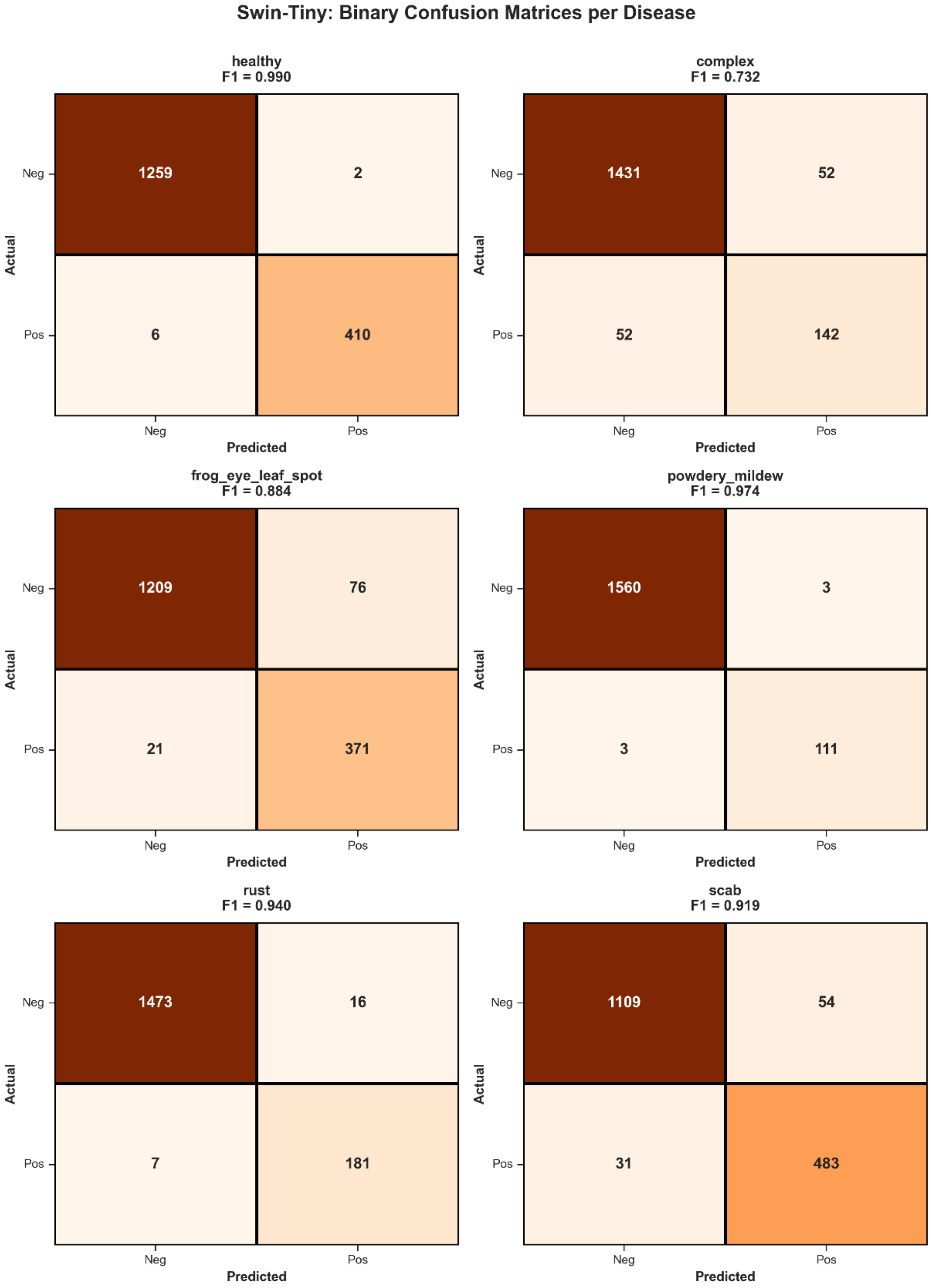


Figure 4.8: Swin-Tiny confusion matrices

4.3.3 ROC-AUC Analysis

ROC curves with AUC values are shown in Figures 4.9 and 4.10. For rust, scab, powdery mildew, and healthy, both models reach AUC≥0.99.

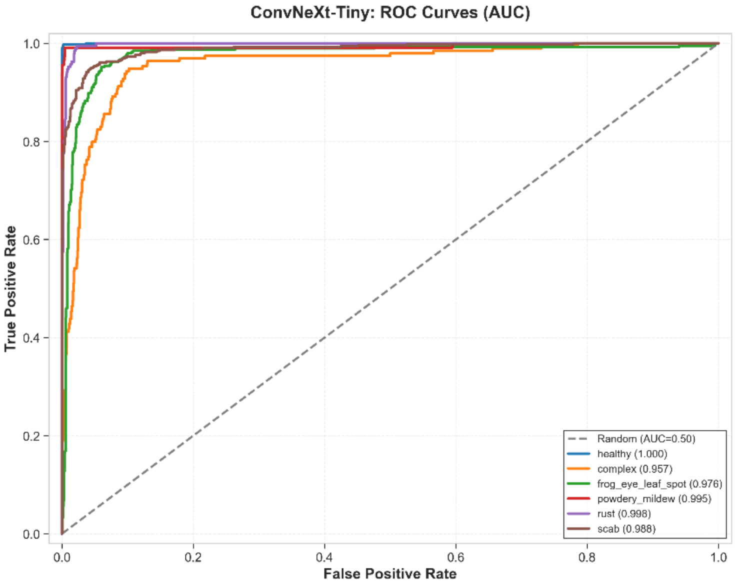


Figure 4.9: ConvNeXt-Tiny ROC curves with AUC values.

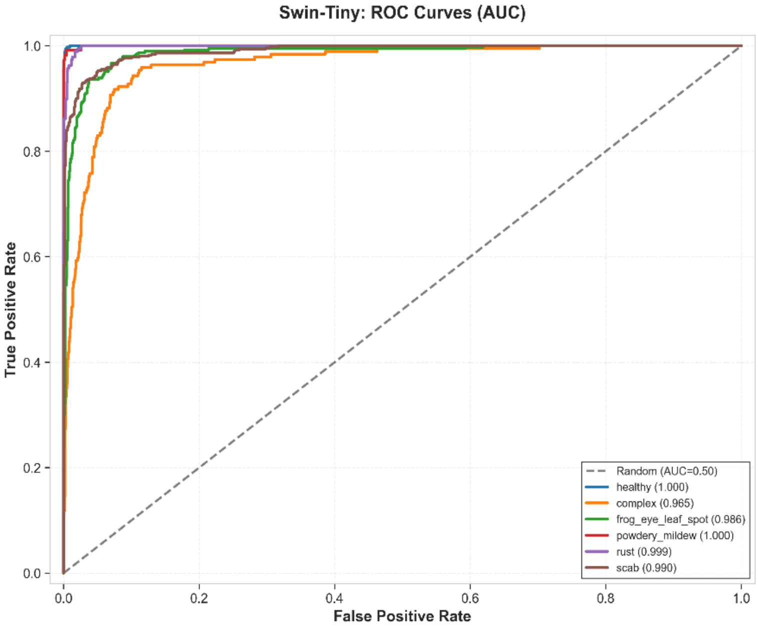


Figure 4.10: Swin-Tiny ROC curves with AUC values.

4.4 Interpretability Analysis

ConvNeXt-Tiny Visualization (Local Focus) in Figure 4.11. Specific dark blemishes on the leaf are sharply and vividly highlighted in this heatmap. This demonstrates how well the CNN architecture detects local features, accurately identifying the precise textures and boundaries of the illness patches.

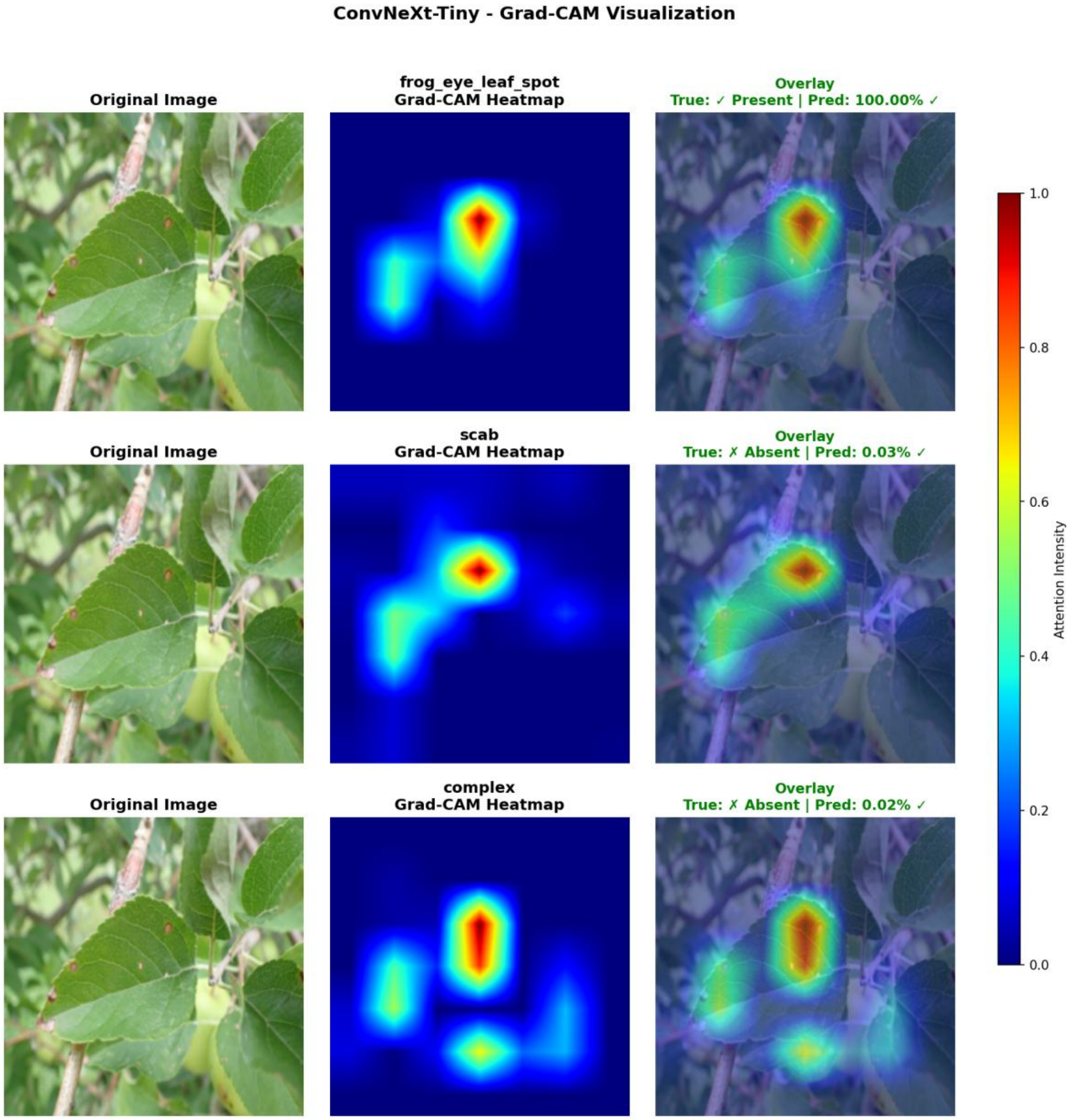


Figure 4.11: ConvNeXt-Tiny Grad-CAM visualization

Swin-Tiny Visualization (Global Focus) in Figure 4.12 large portions of the leaf are covered by this heatmap's dispersed, blocky pattern. This reflects the Swin Transformer's patch-based attention mechanism, which, instead of focusing on specific edges, captures global context and relationships across distant regions of the image.

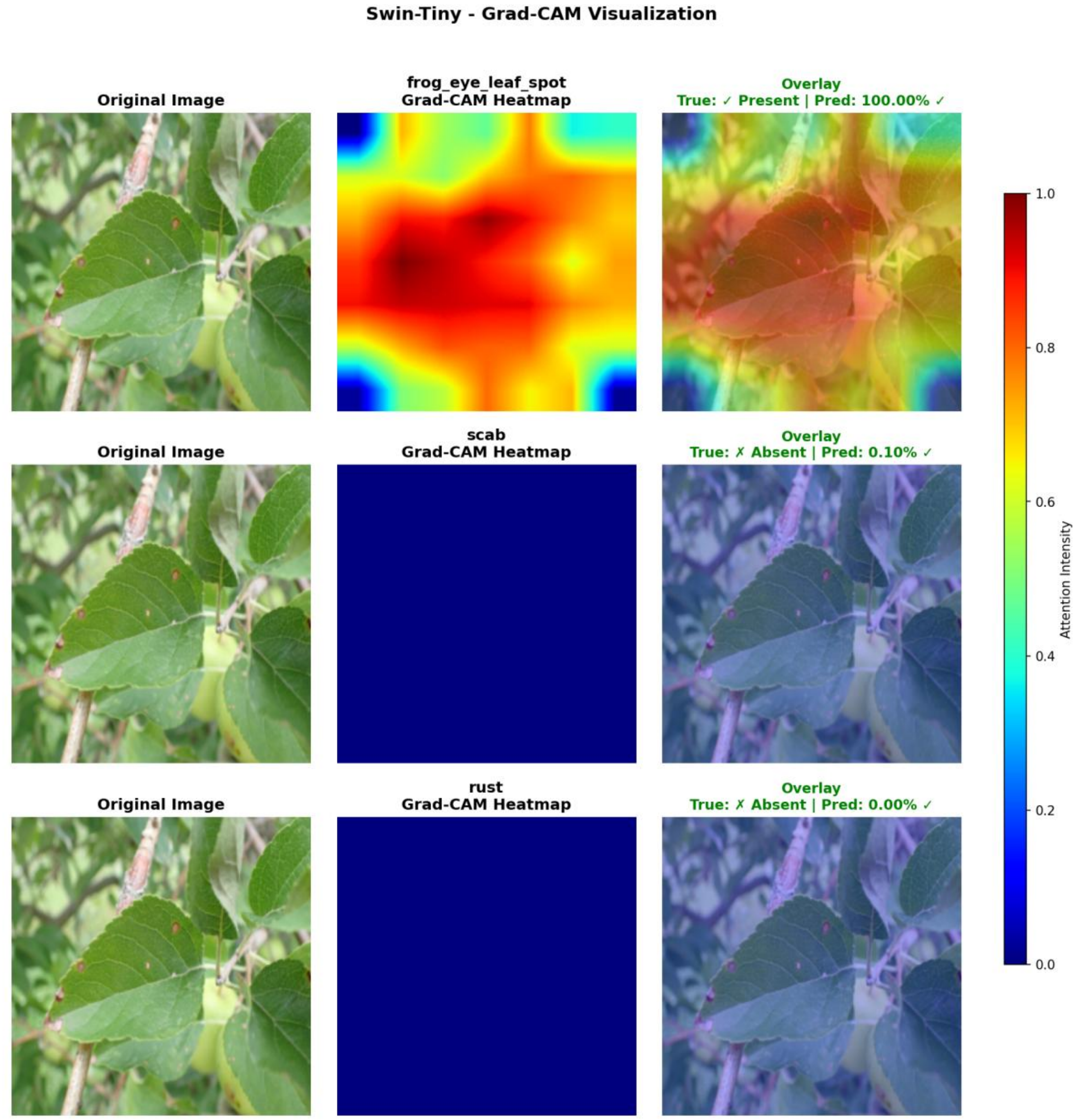


Figure 4.12: Swin-Tiny Grad-CAM visualization

4.5 Discussion

4.5.1 Overall Performance Comparison

The organization achieves its highest performance through its performance-based system which represents its most significant accomplishment. The organization achieves perfect performance because it dedicates all its business activities to delivering results.

ConvNeXt-Tiny has a total performance of 96.90% and 88.61 Exact Match, which shows that it has better performance in comparison to Swin-Tiny. ConvNeXt -Tiny can also be more precise and weighted F1, which shows that it is more resilient to false positives than Swin -Tiny. On the other hand, Swin-Tiny has a small lead in the recall measures, which indicates a relative increase in the sensitivity to detect positive classes. Notably, the Hamming loss of ConvNeXt-Tiny 0.0310 is smaller than Swin -Tiny 0.0321, which means that the former has fewer single label errors per instance in the dataset.

Table 4.3: Performance comparison.

Metric	ConvNeXt-Tiny	Swin-Tiny	Difference
Macro F1	0.9058	0.9066	-0.0008
Micro F1	0.9150	0.9131	+0.0018
Weighted F1	0.9150	0.9136	+0.0014
Macro Precision	0.9001	0.8915	+0.0086
Macro Recall	0.9121	0.9234	-0.0112
Micro Precision	0.9066	0.8932	+0.0134
Micro Recall	0.9235	0.9340	-0.0105
Accuracy	96.90%	96.79%	+0.11%
Exact Match Ratio	88.61%	87.84%	+0.78%
Hamming Loss	0.0310	0.0321	-0.0011

The comparison of the training and validation loss is depicted. It is evident in the plots that both ConvNeXt -Tiny and Swin -Tiny (orange) reach convergence after more than 40 epochs. The convergence in the loss of the training in both the architectures is consistent, with ConvNeXt always recording lower values of loss. By comparison, SwinTiny is more volatile with respect to validation loss, exhibiting jagged spikes, especially at epoch 12, and then levelling off, whereas ConvNeXt is a more even curve.

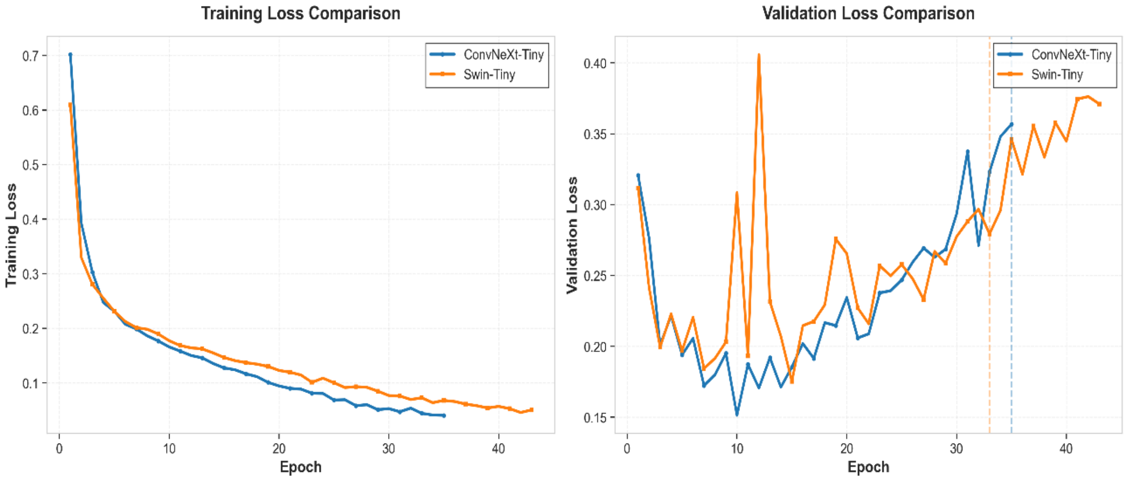


Figure 4.13: Training and Validation Loss Comparison

4.5.2 Per-Disease Performance Comparison

Figure 4.14 shows disease type performance. The findings show fair trade-off between the two architectures: in the case of the healthy, complex, and scab classes, ConvNeXt -Tiny (blue) achieves a little higher F1 scores, but Swin -Tiny (orange) is better at rust, frogeyeleafspot and powdermildew (with the largest improvement of +0.013). The complex class, as compared to other classifications, is a significantly more difficult one regarding both models (score ~0.73).

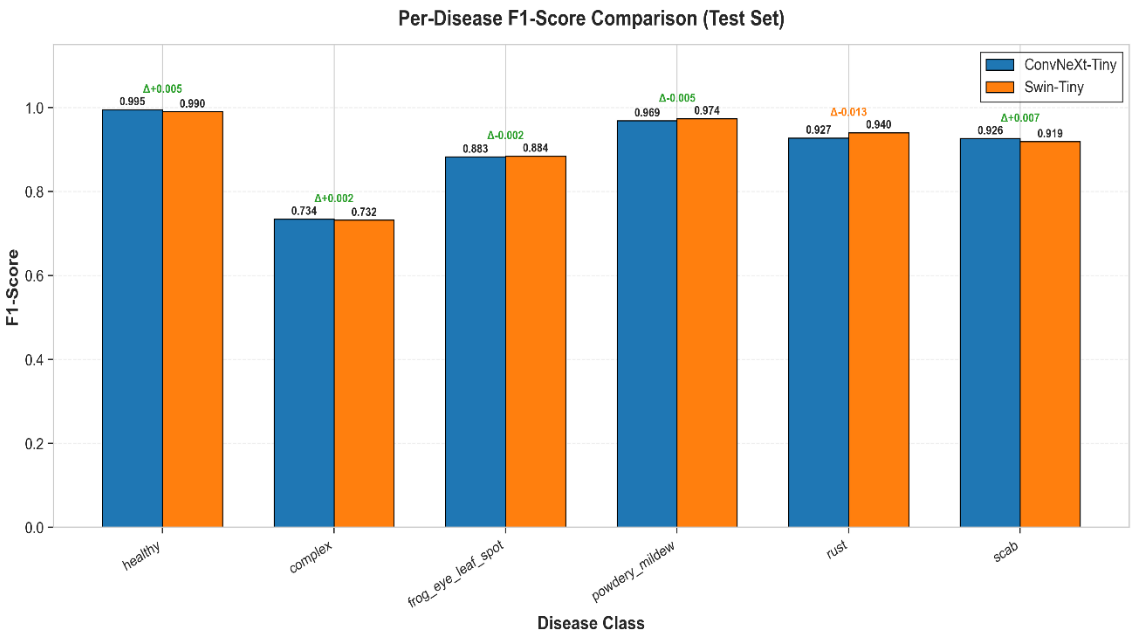


Figure 4.14: Per-disease macro-F1 score comparison

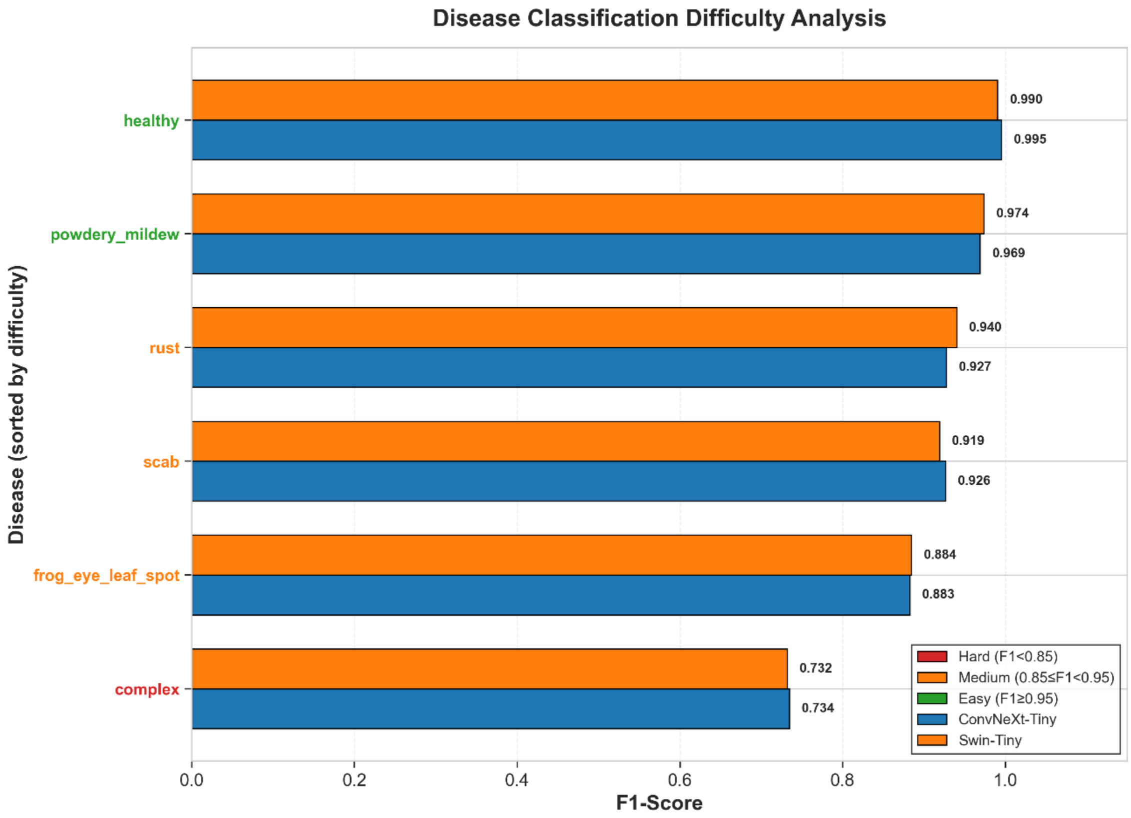


Figure 4.15: Disease classification difficulty analysis

Difficulty Analysis of Disease Classification is given in Figure 4.15. The diseases are categorised as Easy (green), Medium (orange) and Hard (red). The data result in a definite hierarchy: single-disease classes like powdery_mildew and healthy have high reliability (F1> 0.96). The critical bottleneck is the only Hard category with significantly less scores (~0.73) the complex category, which shows that the main difficulty of both algorithms is recognising co-occurring disorders correctly.

4.5.3 Computational Efficiency Comparison

Table 4.4: Computational efficiency comparison.

Metric	**ConvNeXt-Tiny**	**Swin-Tiny**	**Difference**
Total Parameters	27,824,742	**27,523,968**	-300,774
Trainable Parameters	27,824,742	**27,523,968**	-300,774
Model Size	106.14 MB	**105.00 MB**	-1.15 MB
Avg Inference Time (ms/batch)	17.08	**14.29**	-2.79
Inference Time (ms/image)	0.712	**0.595**	-0.116
Throughput (images/sec)	1404.8	**1679.5**	+274.7
Peak GPU Memory	**754.38 MB**	1231.81 MB	+477.43 MB

This analogy brings out a critical trade-off between speed of inference and use of memory. Swin-Tiny operates faster in raw, with a throughput of about 275 images per second and a reduction of inference latency of about 0.12ms per image, which makes it slightly more applicable to high volume, time sensitive operations. ConvNeXt-Tiny, on the other hand, is more memory efficient, at approximately 477MB smaller than Swin-Tiny with regard to GPU memory usage. In terms of storage, both models have similar file sizes (around 105-106 MB) and number of parameters and none of them has a significant advantage in terms of disk space usage.

4.6 Summary

According to the results of the experiment, it is evident that Swin-Tiny shows better performance in terms of sensitivity (recall) and faster inference, whereas, ConvNeXt-Tiny proves to be the more resistant option, as it is more precise, more accurate, and reliable in calibration. The qualitative analysis using Grad-CAM suggests that ConvNeXt is trained on biologically salient lesion-specific features, and Swin uses distributed global attention. The two are also further differentiated by computational benchmarking, which demonstrates that Swin is better suited to high-speed server applications and ConvNeXt to memory-limited edge applications. Regardless of these benefits, there is a performance limitation associated with both of the models in the complex multi-disease group, which highlights the need to conduct future research on the concept of multi-label feature disentanglement. Finally, due to its increased stability, interpretability, and memory efficiency, ConvNeXt-Tiny is recommended as the best architecture to be used in practice and in the field.

CHAPTER 5 CONCLUSION

5.1 Introduction

The current thesis is based on a comparative study of two neural network models, namely the ConvNeXt -Tiny and Swin -Tiny, to perform a multi-label classification of plant diseases, namely in the context of the real-life imagery of apple leaves. By comparing the modern vision models on field-derived photographs of a state of concurrent diseases, the study aims at closing the methodological gap between rigorously controlled laboratory data and the requirements of practical agricultural implementation.

5.2 Key Findings

The two architectures shows very comparable performance metrics with macro F1-scores of 0.9058 and 0.9066 on ConvNeXt and Swin respectively, which is a sign of functional equivalence. Accuracies were above 96, but the exact-match ratios were approximately 88 on each model. The marginal 0.0008 difference in macro F1 does not provide either architecture with statistically or practically significant benefit. System to system precision-recall trade-offs were similar. ConvNeXt-Tiny had a higher precision (0.9001 vs. 0.8915) and a higher exact-match accuracy (88.61 -percent versus 87.84-percent) indicating a conservative high-confidence decision policy. Swin-Tiny on the other hand had a better recall (0.9234 compared to 0.9121) which suggests that it is more sensitive but will detect more false positives. These variations mirror the basic architectural features, that is, locality bias convolutional against global attention mechanisms.

Per-disease analysis was used to indicate a similar ranking of performance in both the models. Exceptional rates of detection were attained with healthy leaves (F1 around 0.99) and powdery mildew (F1 around 0.97) due to the unique visual features. Frogeye leaf spot (F1 breathtakingly close 0.88), rust (F1 breathtakingly close 0.94) and scab (F1 breathtakingly close 0.92) also performed well. Both architectures were the most challenged by the multi-disease composite class (F1^{-} 0.74), which highlights the difficulty of the inherent tasks and not an architectural shortcoming.

Complementary strengths have been corroborated by computational evaluation. Swin-Tiny delivered a 22.4-percent improvement in inference throughput (1,719.9 versus 1,404.8 images per second), thus making it easier to deploy to real time. ConvNeXt -Tiny took 27.6 per cent shorter to train (5.22 hours compared to 6.66 hours), which was helpful in rapid development cycles. Both models had viable memory footprint (less than 1.3GB), which makes it compatible with resource-limited agricultural systems.

The analysis of the Grad-CAM interpretability supported the fact that every architecture learns biologically meaningful features that pay attention to symptoms related to the diseases but not any random background factors. ConvNeXt generated high-resolution local activation maps compared to Swin, which had more localized patterns. The two methods were effective in the classification of disease.

5.3 Research Questions Addressed

- RQ1: Swin-Tiny had a slightly bigger macro F1 (0.9066 as compared to 0.9058); nevertheless, the difference of 0.0008 value indicates functional but not substantive significance.

- RQ2: The performance hierarchies with both models were the same. This was possible even though powdery mildew only made up 6.8 % of data and 0.97 F1, which confirmed the importance of class-weighting approaches. The multi-disease complex that was the most difficult to both architectures (F1≈0.74) indicated inherent task difficulty.
- RQ3: Swin makes a 22.4 per cent faster inference (1,719.9 images per second), and thus allows real time use. ConvNeXt has 27.6% shorter training time (5.22 hrs), which allows the development to develop iteratively. The two models have almost similar memory usage (around 1.2GB) and model size (around 106MB).
- RQ4: Grad-CAM results reveal that ConvNeXt pays attention to local disease characteristics, which is in line with convolutional locality bias, whereas Swin pays attention to larger areas, which is common in global self-attention processes. The two models give importance to biologically relevant symptoms, which is an important validation of meaningful feature learning.

5.4 Limitations of the Research

Dataset: The study is restricted to one crop that is apple and has only six classes of diseases. Unbalanced effects of class imbalance might not be eliminated even when weighted loss functions are used.

Methodology: Only explored the tiny variant of these two model. fixed hyperparameters may not be the best options of both models. The evaluation was made based on one training trial that was not statistically significant.

Computational: Fixed inference conditions; no quantization or hardware acceleration testing.

Evaluation: It was not tested in the field and no study of practitioner user. Only a few samples were analyzed using grad-cam analysis and the temporal disease progression was not assessed.

5.5 Future Implementation of Research

Architectural: Evaluate more variants like Small, Base, Large; compare additional architectures like EfficientNet or other ViT, models then explore ensemble methods combining both models' strengths.

Methodological: Per-disease threshold optimization; semi-supervised learning with unlabeled agricultural data; uncertainty quantification for confidence-aware predictions.

Extension: Use Multi-crop evaluation like tomatoes, grapes, wheat etc. domain adaptation for different imaging conditions; cross-disease transfer learning.

Deployment: It should be deployed through mobile optimisation through quantization and pruning, implementation of of an edge device, and real-time processing pipelines to monitor drones and greenhouses.

Advanced: Temporal modeling of disease progression, multi-modal integration (multispectral, thermal imaging), quantifying the severity of a disease and pixel level segmentation of the image to allow precise treatment are promising research topics.

REFERENCES

Barbedo, J. G. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering, 172*, 84-91. doi:10.1016/j.biosystemseng.2018.05.013

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*. Retrieved from https://arxiv.org/abs/2010.11929

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). Las Vegas, Nevada: IEEE . doi:10.1109/CVPR.2016.90

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., . . . Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint, arXiv:1704.04861*. Retrieved from https://arxiv.org/abs/1704.04861

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4700-4708). Honolulu, Hawaii: IEEE. doi:10.1109/CVPR.2017.243

Jiang, P., Chen, Y., He, D., & Liang, C. (2020). Deep Learning-Based Detection of Apple Leaf Diseases Using Images Taken in Field Conditions. *Plant Disease, 104*(8), 2099-2108. doi:10.1094/PDIS-11-19-2445-FE

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70-90. doi:10.1016/j.compag.2018.02.016

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25* (pp. 1097-1105). Lake Tahoe: Curran Associates, Inc.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2117-2125). Honolulu, Hawaii, USA.: IEEE .

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 2980-2988). Venice, Italy: IEEE.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., . . . Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. *arXiv preprint, arXiv:2103.14030*. Retrieved from https://arxiv.org/abs/2103.14030

Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. *arXiv preprint, arXiv:2201.03545*. Retrieved from https://arxiv.org/abs/2201.03545

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. *International Conference on Learning Representations (ICLR).* New Orleans: ICLR. Retrieved from https://arxiv.org/abs/1711.05101

Maree, H. J., Fox, A., Al Rwahnih, M., Boonham, N., & Candresse, T. (2018). Application of HTS for routine plant virus diagnostics: State of the art and challenges. *Frontiers in Plant Science, 9*, 1082. doi:10.3389/fpls.2018.01082

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, 1419. doi:10.3389/fpls.2016.01419

Mottet, A., de Haan, C., Falcucci, A., Tempio, G., Opio, C., & Gerber, P. (2017). Livestock: On our plates or eating at our table? A new analysis of the feed/food debate. *Global Food Security, 14*, 1-8. doi:10.1016/j.gfs.2017.01.001

Müller, F., Jaeger, D., & Hanewinkel, M. (2019). Digitization in wood supply – A review on how Industry 4.0 will change the forest value chain. *Computers and Electronics in Agriculture*, 206-218. doi:10.1016/j.compag.2019.04.002

Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants, 8*(11), 468. doi:10.3390/plants8110468

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution, 3*(3), 430-439. doi:10.1038/s41559-018-0793-y

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 618-626). Venice, Italy: IEEE.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning (ICML)* (pp. 6105-6114). Long Beach, California, USA.: Proceedings of Machine Learning Research. Retrieved from https://arxiv.org/abs/1905.11946

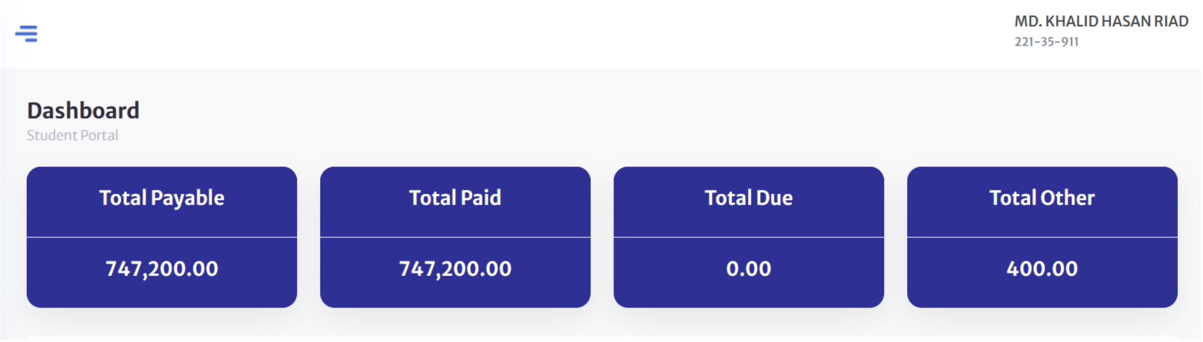
Thapa, R., Zhang, K., Snavely, N., Belongie, S., & Khan, A. (2020). The Plant Pathology Challenge 2020 data set to classify foliar disease of apples. *Applications in Plant Sciences, 8*(9), e11390. doi:10.1002/aps3.11390

Thapa, R., Zhang, K., Snavely, N., Belongie, S., & Khan, A. (2021). *Plant Pathology 2021 - FGVC8*. Retrieved from Kaggle: https://kaggle.com/competitions/plant-pathology-2021-fgvc8

Wightman, R. (2019). *PyTorch Image Models*. Retrieved from GitHub: https://github.com/rwightman/pytorch-image-models

Zhang, M.-L., & Zhou, Z.-H. (2014). A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering, 26*(8), 1819-1837. doi:10.1109/TKDE.2013.39

Account Clearance

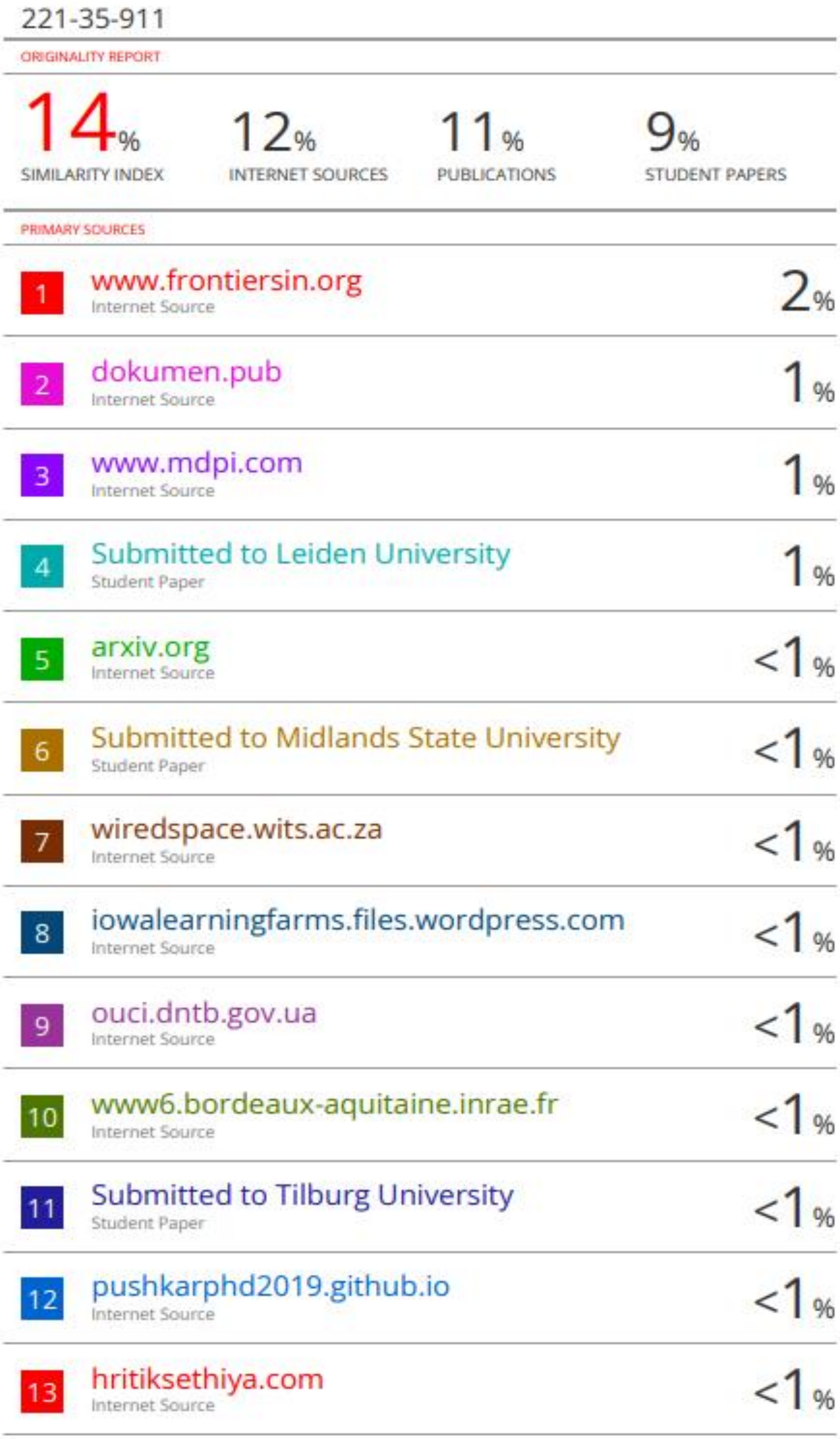


MD. KHALID HASAN RIAD
221-35-911

Dashboard
Student Portal

Total Payable	Total Paid	Total Due	Total Other
747,200.00	747,200.00	0.00	400.00

Plagiarism Report



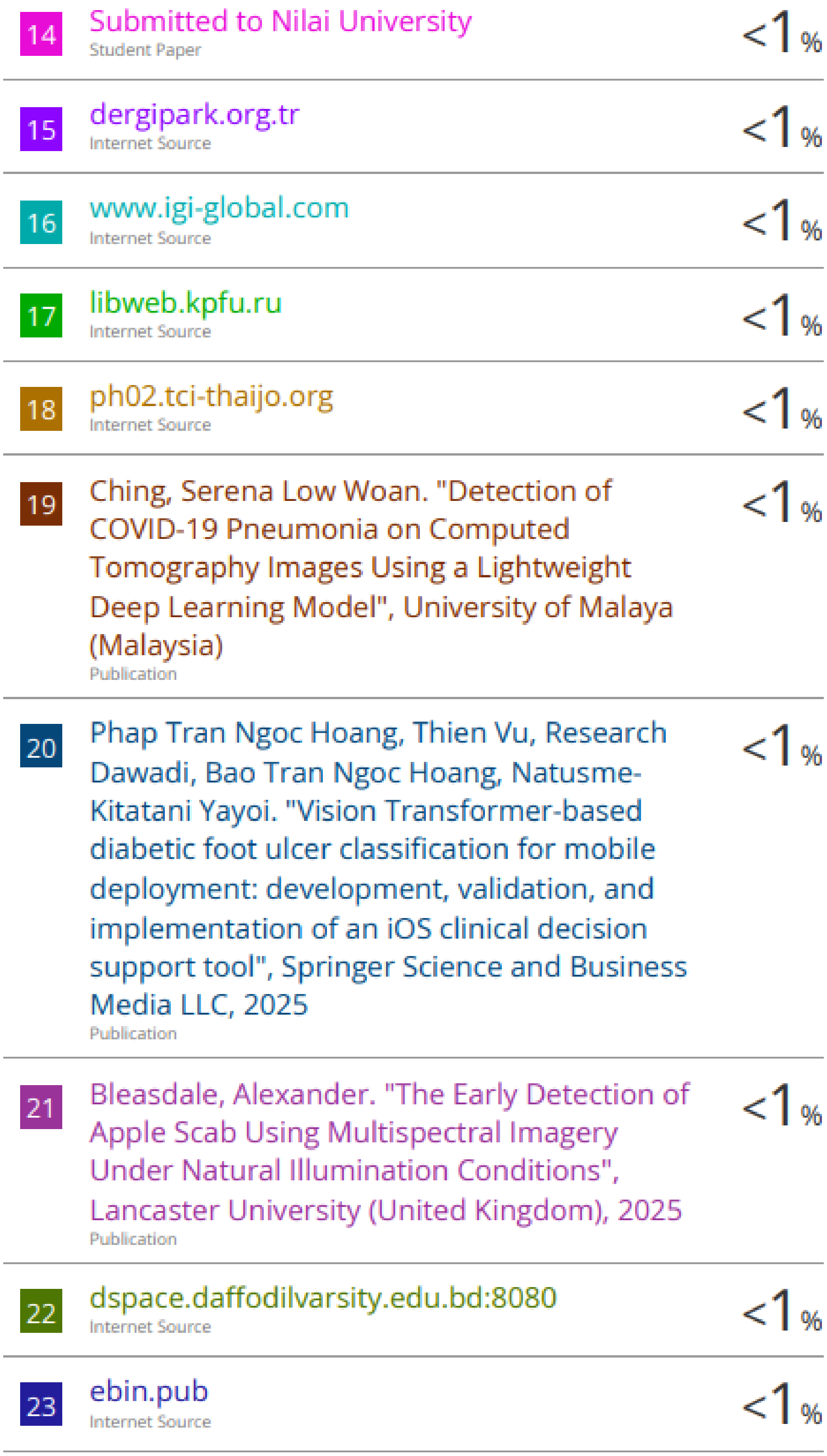
221-35-911

ORIGINALITY REPORT

14%	12%	11%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

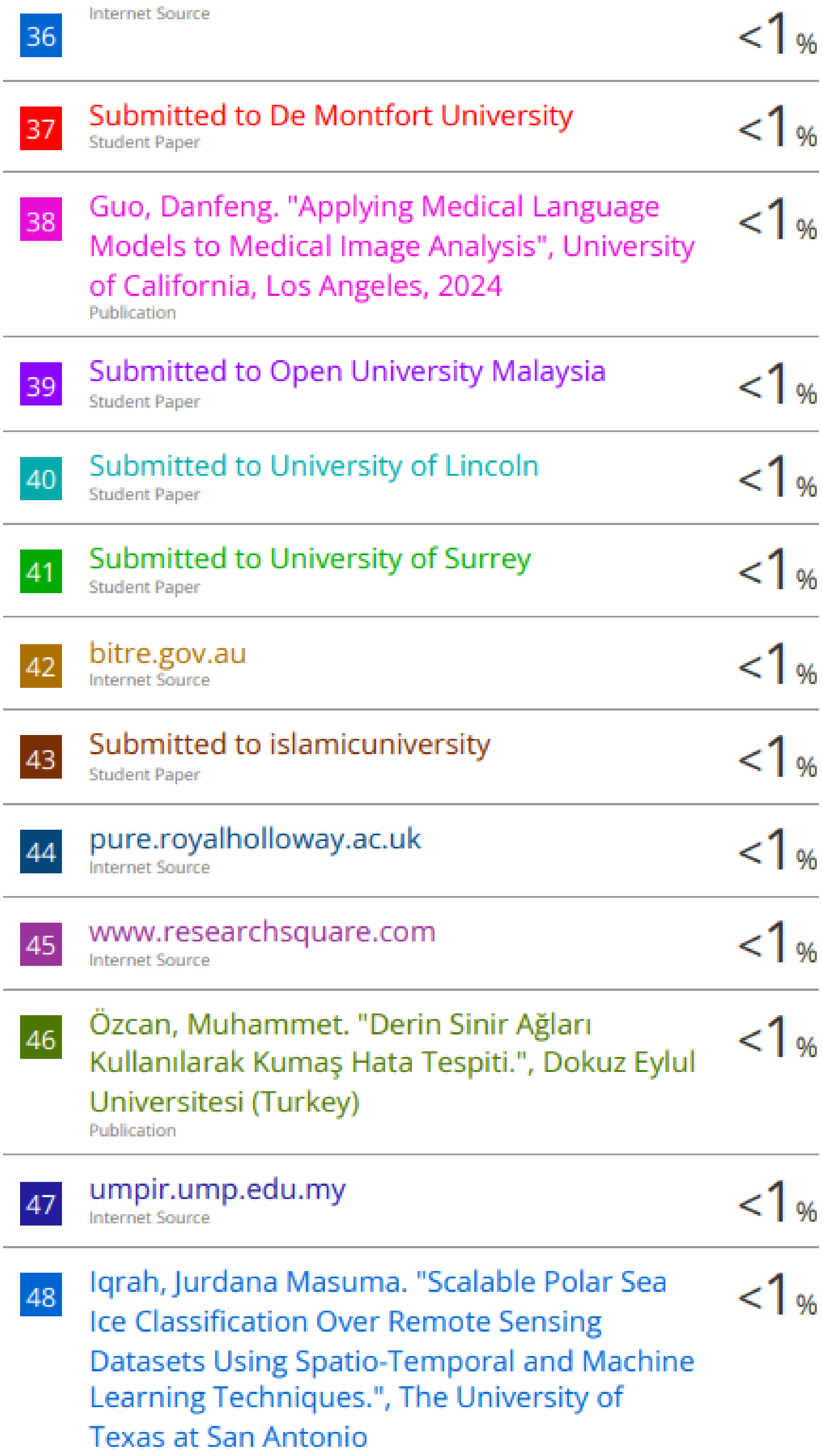
PRIMARY SOURCES

1	www.frontiersin.org Internet Source	2%
2	dokumen.pub Internet Source	1%
3	www.mdpi.com Internet Source	1%
4	Submitted to Leiden University Student Paper	1%
5	arxiv.org Internet Source	<1%
6	Submitted to Midlands State University Student Paper	<1%
7	wiredspace.wits.ac.za Internet Source	<1%
8	iowalearningfarms.files.wordpress.com Internet Source	<1%
9	ouci.dntb.gov.ua Internet Source	<1%
10	www6.bordeaux-aquitaine.inrae.fr Internet Source	<1%
11	Submitted to Tilburg University Student Paper	<1%
12	pushkarphd2019.github.io Internet Source	<1%
13	hritiksethiya.com Internet Source	<1%



14	Submitted to Nilai University Student Paper	<1 %
15	dergipark.org.tr Internet Source	<1 %
16	www.igi-global.com Internet Source	<1 %
17	libweb.kpfu.ru Internet Source	<1 %
18	ph02.tci-thaijo.org Internet Source	<1 %
19	Ching, Serena Low Woan. "Detection of COVID-19 Pneumonia on Computed Tomography Images Using a Lightweight Deep Learning Model", University of Malaya (Malaysia) Publication	<1 %
20	Phap Tran Ngoc Hoang, Thien Vu, Research Dawadi, Bao Tran Ngoc Hoang, Natusme-Kitatani Yayoi. "Vision Transformer-based diabetic foot ulcer classification for mobile deployment: development, validation, and implementation of an iOS clinical decision support tool", Springer Science and Business Media LLC, 2025 Publication	<1 %
21	Bleasdale, Alexander. "The Early Detection of Apple Scab Using Multispectral Imagery Under Natural Illumination Conditions", Lancaster University (United Kingdom), 2025 Publication	<1 %
22	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1 %
23	ebin.pub Internet Source	<1 %

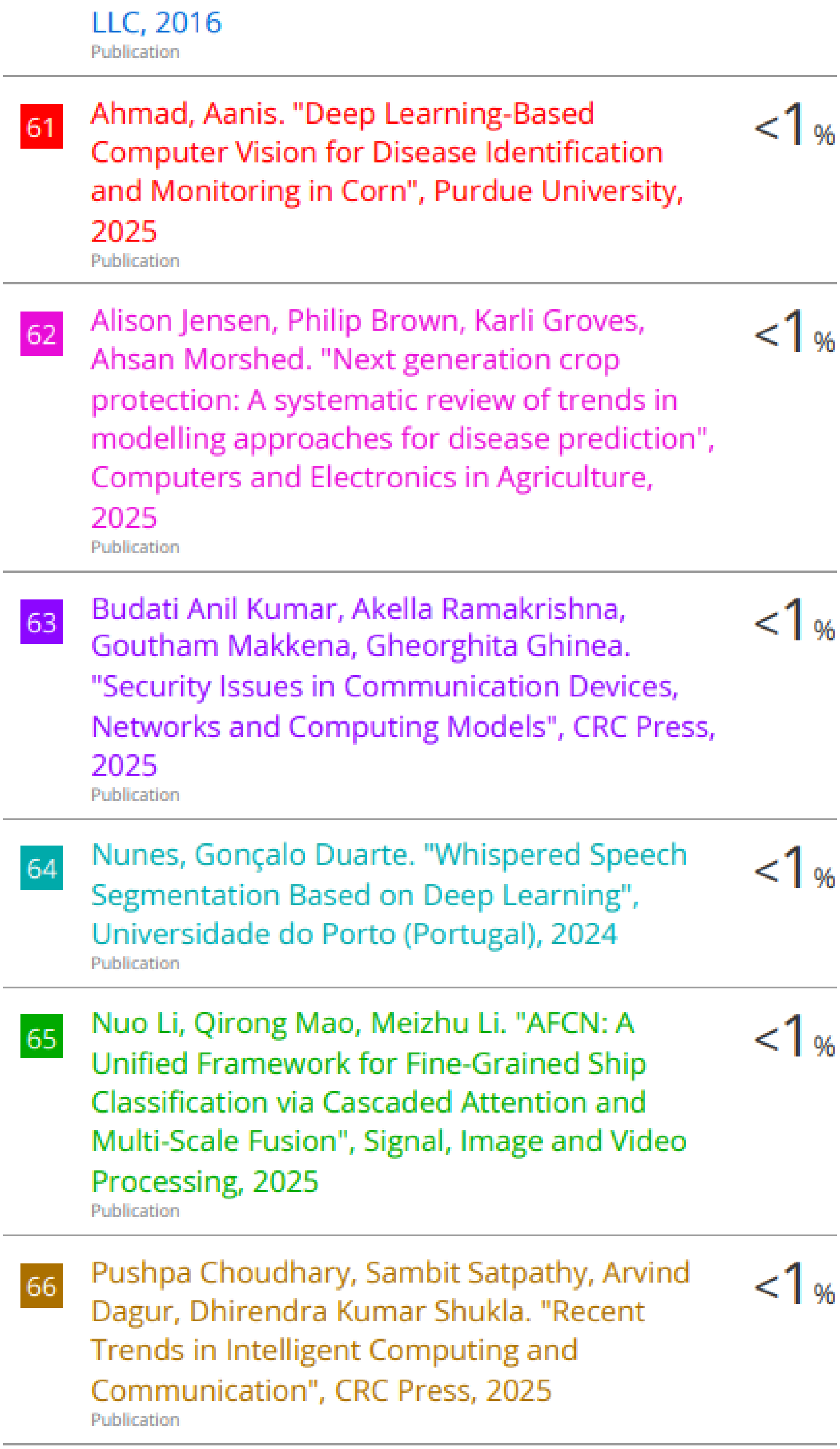
24	dspace.ewubd.edu:8080 Internet Source	<1 %
25	Submitted to University of Reading Student Paper	<1 %
26	academic.oup.com Internet Source	<1 %
27	Submitted to The University of Manchester Student Paper	<1 %
28	marswiki.jrc.ec.europa.eu Internet Source	<1 %
29	Submitted to Higher Education Commission Pakistan Student Paper	<1 %
30	S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufacturing", CRC Press, 2025 Publication	<1 %
31	Simon C.K. Chan, Bingxin Huang, Hannah H. Kim, Victor T.C. Tsang, Terence T.W. Wong. "Wavelet-Enhanced Residual Optimal Transport for Mamba-Based Image Restoration in Photoacoustic Tomography", Photoacoustics, 2025 Publication	<1 %
32	mro.massey.ac.nz Internet Source	<1 %
33	Submitted to National University of Singapore Student Paper	<1 %
34	Submitted to University of Queensland Student Paper	<1 %
35	cdr.lib.unc.edu Internet Source	<1 %
	studentsrepo.um.edu.my	



36	Internet Source	<1 %
37	Submitted to De Montfort University Student Paper	<1 %
38	Guo, Danfeng. "Applying Medical Language Models to Medical Image Analysis", University of California, Los Angeles, 2024 Publication	<1 %
39	Submitted to Open University Malaysia Student Paper	<1 %
40	Submitted to University of Lincoln Student Paper	<1 %
41	Submitted to University of Surrey Student Paper	<1 %
42	bitre.gov.au Internet Source	<1 %
43	Submitted to islamicuniversity Student Paper	<1 %
44	pure.royalholloway.ac.uk Internet Source	<1 %
45	www.researchsquare.com Internet Source	<1 %
46	Özcan, Muhammet. "Derin Sinir Ağları Kullanılarak Kumaş Hata Tespiti.", Dokuz Eylul Universitesi (Turkey) Publication	<1 %
47	umpir.ump.edu.my Internet Source	<1 %
48	Iqrah, Jurdana Masuma. "Scalable Polar Sea Ice Classification Over Remote Sensing Datasets Using Spatio-Temporal and Machine Learning Techniques.", The University of Texas at San Antonio	<1 %

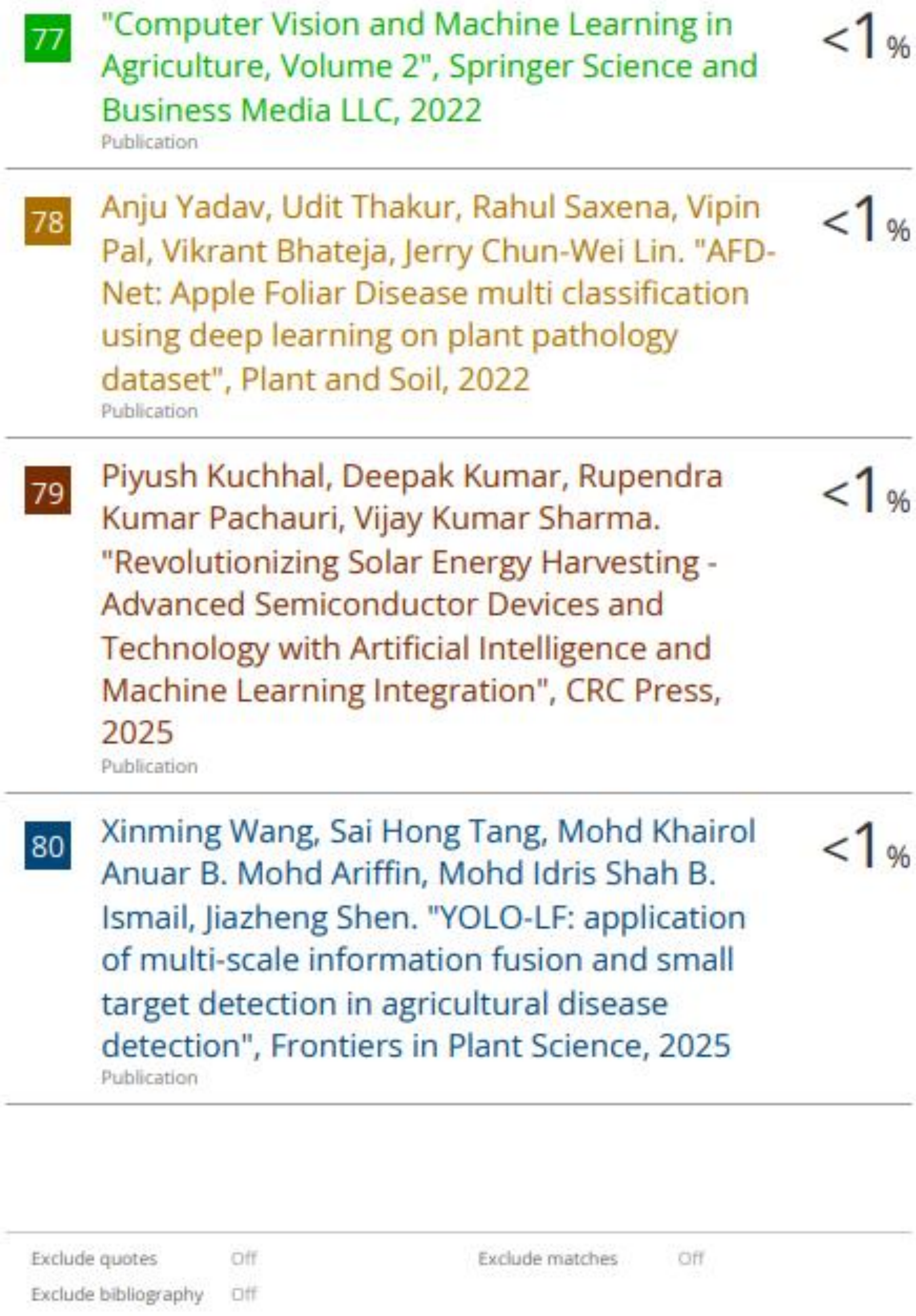
Publication

49	Jianping Yao, Son N. Tran, Samantha Sawyer, Saurabh Garg. "Machine learning for leaf disease classification: data, techniques and applications", Artificial Intelligence Review, 2023 Publication	<1 %
50	dspace.cvut.cz Internet Source	<1 %
51	jfde.eu Internet Source	<1 %
52	peltarion.com Internet Source	<1 %
53	www.atyun.com Internet Source	<1 %
54	www.medrxiv.org Internet Source	<1 %
55	www.preprints.org Internet Source	<1 %
56	7universum.com Internet Source	<1 %
57	Md. Nurullah, Rania Hodhod, Hyrum Carroll, Yi Zhou. "Advancing Multi-Label Tomato Leaf Disease Identification Using Vision Transformer and EfficientNet with Explainable AI Techniques", Electronics, 2025 Publication	<1 %
58	eprints.uthm.edu.my Internet Source	<1 %
59	www.hindawi.com Internet Source	<1 %
60	"Advances in Artificial Intelligence - IBERAMIA 2016", Springer Science and Business Media	<1 %



	LLC, 2016 Publication	
61	Ahmad, Aanis. "Deep Learning-Based Computer Vision for Disease Identification and Monitoring in Corn", Purdue University, 2025 Publication	<1 %
62	Alison Jensen, Philip Brown, Karli Groves, Ahsan Morshed. "Next generation crop protection: A systematic review of trends in modelling approaches for disease prediction", Computers and Electronics in Agriculture, 2025 Publication	<1 %
63	Budati Anil Kumar, Akella Ramakrishna, Goutham Makkena, Gheorghita Ghinea. "Security Issues in Communication Devices, Networks and Computing Models", CRC Press, 2025 Publication	<1 %
64	Nunes, Gonçalo Duarte. "Whispered Speech Segmentation Based on Deep Learning", Universidade do Porto (Portugal), 2024 Publication	<1 %
65	Nuo Li, Qirong Mao, Meizhu Li. "AFCN: A Unified Framework for Fine-Grained Ship Classification via Cascaded Attention and Multi-Scale Fusion", Signal, Image and Video Processing, 2025 Publication	<1 %
66	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1 %

67	Suneeta Satpathy, Álvaro Rocha, Sachi Nandan Mohanty, Tanupriya Choudhury. "Intelligent Data-Driven Systems with Innovations in Artificial Intelligence", CRC Press, 2025 Publication	<1 %
68	Submitted to Universiteit Twente Student Paper	<1 %
69	core.ac.uk Internet Source	<1 %
70	indah.ump.edu.my Internet Source	<1 %
71	link.springer.com Internet Source	<1 %
72	publications.polymtl.ca Internet Source	<1 %
73	ul.netd.ac.za Internet Source	<1 %
74	Guimarães, Vânia Margarida Cardoso. "Machine Learning and Movement Analysis for Cognitive Screening in Older Adults", Universidade do Porto (Portugal), 2025 Publication	<1 %
75	Sivasubramaniam Janarthan, Selvarajah Thuseethan, Sutharshan Rajasegarar, John Yearwood. "P2OP—Plant Pathology on Palms: A deep learning-based mobile solution for in-field plant disease detection", Computers and Electronics in Agriculture, 2022 Publication	<1 %
76	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	<1 %



77	"Computer Vision and Machine Learning in Agriculture, Volume 2", Springer Science and Business Media LLC, 2022 Publication	<1 %
78	Anju Yadav, Udit Thakur, Rahul Saxena, Vipin Pal, Vikrant Bhateja, Jerry Chun-Wei Lin. "AFD-Net: Apple Foliar Disease multi classification using deep learning on plant pathology dataset", Plant and Soil, 2022 Publication	<1 %
79	Piyush Kuchhal, Deepak Kumar, Rupendra Kumar Pachauri, Vijay Kumar Sharma. "Revolutionizing Solar Energy Harvesting - Advanced Semiconductor Devices and Technology with Artificial Intelligence and Machine Learning Integration", CRC Press, 2025 Publication	<1 %
80	Xinming Wang, Sai Hong Tang, Mohd Khairol Anuar B. Mohd Ariffin, Mohd Idris Shah B. Ismail, Jiazheng Shen. "YOLO-LF: application of multi-scale information fusion and small target detection in agricultural disease detection", Frontiers in Plant Science, 2025 Publication	<1 %

Exclude quotes Off
Exclude bibliography Off
Exclude matches Off