

An Empirical Study with Pre Trained and Attention Model on Food Disease Classification

*A Project submitted in partial fulfillment of requirements
For the award of the degree of
Master of Computer Application*

by

Sohom Roy Choudhury

Regn. No.- 221040510049 Exam Roll No. - 12022010010038

Rishav Bhattacharjee

Regn. No.- 221040510036 Exam Roll No. - 12022010010008

Ayan Batabyal

Regn. No.- 221040510011 Exam Roll No. - 12022010010003

under the supervision of

Prof. Supratim Ghosh

Assistant Professor

&

Prof. Dr. Priti Deb

Assistant Professor

Department of Computer Application
INSTITUTE OF ENGINEERING & MANAGEMENT
Kolkata, West Bengal, India
2022-2024

Declaration Certificate

This is to certify that the work presented in the thesis entitled "**An Empirical Study with Pre-Trained and Attention Model on Food Disease Classification**" in partial fulfilment of the requirement for the award of degree of **Master of Computer Application** of Institute of Engineering & Management is an authentic carried out under my supervision and guidance.

To the best of my knowledge the content of this thesis dose not form a basis for the award of any previous degree to anyone else.

Prof. Supratim Ghosh,
Assistant Professor,
Department of Computer Application,
Institute of Engineering & Managemnet.
(Supervisor)

Prof. Dr. Priti Deb,
Assistant Professor,
Department of Computer Application,
Institute of Engineering & Managemnet.
(Supervisor)

Head of the Department,
Master of Computer Application and Sciences,
Institute of Engineering & Management.

Certificate of Approval

The for going thesis entitled "**An Empirical Study with Pre-Trained and Attention Model on Food Disease Classification**" is hereby approved as a creditable study of research topic and has been presented in satisfactory manner to warrant its acceptance as prerequisite to the degree for which it has been submitted.

It is understood by this approval, the undersigned do not necessarily endorse any conclusion or opinion expressed there in, but approve the thesis for the purpose for which it is submitted.

Examiners:

(Signature of The Examiner)

(Signature of The Supervisor)

(Signature of The Supervisor)

ACKNOWLEDGEMENT

We would like to express our special thanks to our supervisor, Prof. Supratim Ghosh and Prof. Dr. Priti Deb who helped us a lot in this project, their valuable suggestions helped us to solve tough challenges and without their help this project could not have been completed in time. A special thanks to them who gave us the golden opportunity to do this wonderful project on the topic “An Empirical Study with Pre Trained and Attention Model on Food Disease Classification”, which helped us to gain a significant knowledge in the aforesaid subjects. We would also like to express our warm regards as a note of thanks to Prof. Dr. Pawan Kumar Singh of Jadavpur University(IT Department) and Prof. Debam Saha of Calcutta Institute of Engineering and Management(CSE Department) for their wonderful support for their help on this topic. Secondly, we would like to thank our friends who helped us a lot in finalising this project within the given time frame.

Sohom Roy Choudhury

Roll No. - 12022010010038

Department of Computer Application,
Institute of Engineering & Management.

Rishav Bhattacharjee

Roll No. - 12022010010008

Department of Computer Application,
Institute of Engineering & Management.

Ayan Batabyal

Roll No. - 12022010010003

Department of Computer Application,
Institute of Engineering & Management.

Contents

Certificate from Supervisor	I
Certificate of Approval	II
Acknowledgement	III
1 Introduction	1
1.1 Research Motivation	2
1.2 Organization of Thesis	2
2 Literature Survey	4
3 Methodology	7
3.1 Materials and methods	7
3.1.1 VGG	8
3.1.2 ResNet	9
3.1.3 DenseNet	10
3.1.4 MobileNet	12
3.1.5 Squeeze-Excitation Attention Networks	13
4 Experiments	15
4.1 Hardware Used	15
4.2 Optimizer	16
4.3 Loss Function	16
5 Dataset Details	17
5.1 PLD(Potato Leaf Disease) dataset	17
5.2 Fruit Infection Disease dataset	18
5.3 Dataset Statistics	21
6 Result and Analysis	23
6.1 Evaluation	23
6.2 Classification metrics	23
6.3 Confusion Matrix	24
6.4 Experimental Result	24

6.5	Performance Comparison between Top-3 State of the Art for the Dataset PLD (Potato Leaf Disease)	25
6.6	Performance Comparison between Top-3 State of the Art with Attention Model for the Dataset PLD (Potato Leaf Disease)	26
6.7	Performance Comparison between Top-3 State of the Art for the Dataset FID (Fruit Infection Disease)	27
6.8	Performance Comparison between Top-3 State of the Art with Attention Model for the Dataset FID (Fruit Infection Disease)	28
6.9	Discussion	29
7	Conclusion and Future Work	30
7.1	Conclusion	30
7.2	Future Work	30

List of Figures

3.1	VGG Architecture	9
3.2	ResNet Architecture	10
3.3	DenseNet Architecture	12
3.4	MobileNet Architecture	13
3.5	Squeeze - Excitation Attention	14
5.1	Sample Images from PLD(Potato Leaf Disease) dataset	18
5.2	Sample Images from Strawberry Anthracnos Fruit Rot	19
5.3	Sample Images from Strawberry Gray Mold	19
5.4	Sample Images from Tomato Leaf Mold	19
5.5	Sample Images from Tomato Spider Miles	20
5.6	Sample Images from Beans Rust	20
5.7	Sample Images from Beans Angular Leaf Spot	20
5.8	Bar graph depicting PLD dataset distribution	21
5.9	Bar graph depicting FID dataset distribution	22
6.1	Confusion Matrix Achieved on PLD using Only Base model DenseNet121	25
6.2	Confusion Matrix Achieved on PLD using Only Base model MobileNet	25
6.3	Confusion Matrix Achieved on PLD using Only Base model ResNet101V2	25
6.4	Confusion Matrix Achieved on PLD using Base model DenseNet121 with SE Attention Block	26
6.5	Confusion Matrix Achieved on PLD using Base model MobileNet with SE Attention Block	26
6.6	Confusion Matrix Achieved on PLD using Base model ResNet101V2 with SE Attention Block	26
6.7	Confusion Matrix Achieved on FID using Only Base model MobileNet	27
6.8	Confusion Matrix Achieved on FID using Only Base model ResNet101V2	27
6.9	Confusion Matrix Achieved on FID using Only Base model DenseNet121	27
6.10	Confusion Matrix Achieved on FID using Base model with SE Attention Block ResNet101V2	28
6.11	Confusion Matrix Achieved on FID using Base model with SE Attention Block MobileNet	28
6.12	Confusion Matrix Achieved on FID using Base model with SE Attention Block DenseNet121	28

List of Tables

6.1	Top-3 State of the art for the Dataset PLD	25
6.2	Top-3 State of the art for the Dataset PLD with Attention Model	26
6.3	Top-3 State of the art for the Dataset FID	27
6.4	Top-3 State of the art for the Dataset FID with Attention Model	28

Abstract

Food is an important sector for a country which gets affected by an innumerable diseases at an early stage. These diseases affect the various crops which ultimately decreases the productivity of these food crops. This thesis investigates the classification of food diseases employing deep learning models. Various well-established architectures such as DenseNet201, DenseNet121, ResNet50V2, ResNet101V2, VGG16, VGG19, MobileNetV2, and MobileNet are trained on two distinct datasets: Potato Leaf Disease (PLD) and Food Infection Disease (FID). Notably, among these models, MobileNet, ResNet101V2, and DenseNet121 exhibit superior performance. Furthermore, the integration of an attention mechanism, specifically the squeeze and excitation (SE) module, enhances classification accuracy significantly. This augmentation serves as a pivotal motivation for the present study. By incorporating attention layers, the thesis demonstrates a notable improvement in accuracy levels, underscoring the potential of attention mechanisms in deep learning-based disease classification tasks. Such findings not only advance the understanding of deep learning applications in disease diagnosis but also highlight the importance of attention mechanisms in refining model performance.

Keyword: Deep Learning, Food Disease Classification, Squeeze and Excitation, DenseNet201, DenseNet121, ResNet50V2, ResNet101V2, VGG16, VGG19, MobileNetV2, MobileNet, Potato Leaf Disease (PLD) and Food Infection Disease (FID).

Chapter 1

Introduction

The exploration of deep learning methodologies in revolutionizing disease classification within the agricultural staples of potatoes and apples represents a critical endeavor in addressing the pervasive threats posed by crop diseases. These two staples, integral to global sustenance and economies, are constantly besieged by a myriad of diseases orchestrated by insidious invaders such as fungi, bacteria, and viruses. The repercussions of these diseases reverberate far beyond individual farms, exerting profound impacts on the livelihoods of farmers and the broader global food security landscape. Thus, the timely and precise detection of diseases stands as a pivotal linchpin for implementing targeted interventions, minimizing yield losses, and safeguarding food supplies.

Traditionally, the laborious task of disease detection has relied on visual inspections conducted by experts. However, this method is fraught with challenges—it is time-consuming, labor-intensive, and inherently susceptible to human error. Recognizing the limitations inherent in traditional approaches, this research heralds a paradigm shift by harnessing the capabilities of deep learning to surmount these challenges. Eight distinct deep learning models, each endowed with unique strengths, are slated for rigorous evaluation. Among them are DenseNet201 and ResNet50V2, celebrated for their prowess in image classification, alongside MobileNetV2, tailored for efficiency on mobile devices. The primary objective is to assess their efficacy in accurately classifying a broad spectrum of potato leaf and apple diseases. This comprehensive evaluation will be multifaceted, scrutinizing classification accuracy, computational efficiency, and scalability. The overarching aim is to pinpoint the front runner—the model capable of delivering superior disease classification accuracy without exacting an exorbitant computational toll.

However, the research endeavor does not culminate with model evaluation. It delves deeper into the realm of transfer learning and fine-tuning strategies to explore their potential in further enhancing model performance. Consider a deep learning model pre-trained on an extensive dataset of generic plant images. Such a model has already assimilated a wealth of knowledge about visual features. Transfer learning empowers researchers to leverage this pre-existing knowledge and adapt it to the specific challenge of potato and apple disease classification. Fine-tuning takes this adaptation a step further, meticulously calibrating the model's internal parameters to refine its ability to discern healthy crops from those afflicted by disease. This exploration holds promise not only for augmenting model performance but also for significantly curtailing the time required for training.

Ultimately, the research endeavors to bridge the chasm between cutting-edge technological innovations and the pragmatic exigencies of real-world agricultural settings. By rigorously evaluating

the performance of these deep learning models, the research aims to ascertain their viability as reliable and efficient tools for farmers. Envision a future where farmers can seamlessly leverage a smartphone application equipped with a deep learning model to swiftly and accurately diagnose diseases in their crops. Such technology harbors the potential to transform crop health monitoring, enabling early detection and intervention and thereby fostering a more resilient food supply chain and sustainable agricultural practices.

Moreover, the implications of this research extend beyond disease classification alone. By democratizing access to advanced technological solutions, such as deep learning models, even to resource-constrained farming communities, the research contributes to narrowing the digital divide in agriculture. Furthermore, the integration of deep learning into agricultural practices opens avenues for data-driven decision-making, precision agriculture, and the optimization of resource allocation. By harnessing the power of data analytics and artificial intelligence, farmers can optimize inputs, minimize waste, and maximize yields, thereby fostering greater sustainability in agricultural production.

In essence, the transformative potential of deep learning in revolutionizing disease classification within potatoes and apples signifies a watershed moment in agricultural innovation. By transcending the limitations of traditional disease detection methods, deep learning empowers farmers with the tools needed to safeguard their crops, enhance productivity, and fortify global food security. As such, this research not only underscores the pivotal role of technology in addressing contemporary agricultural challenges but also heralds a future where innovation serves as a cornerstone for building resilient and sustainable food systems.

1.1 Research Motivation

Agriculture is a cornerstone of South Asian economies, including India. This thesis focuses on a critical challenge: early and accurate detection of crop diseases like Potato Leaf Disease (PLD) and others causing Food Infection Diseases (FID). Early detection minimizes crop losses and wasted human labor, a crucial resource in the region. This research proposes a solution using deep learning to automate disease classification. By automating this process, farmers can save valuable time and effort traditionally spent on manual inspection. Additionally, accurate classification ensures only healthy crops reach consumers. This translates to better food quality and contributes to improved public health and hygiene. In essence, this thesis aims to leverage technology for a two-fold benefit: protecting crops and enhancing public health throughout South Asia.

1.2 Organization of Thesis

The thesis is distributed into seven chapters. Chapter 1 provides an introduction to the thesis. The rest of the thesis is as follows:

- **Chapter 2** provides a literature review of several classification-based approaches. In this chapter, we have reviewed the publications that have been conducted on the topic of categorization.
- **Chapter 3** provides information about the methodology of our work. In this chapter, we have provided information about all the state-of-the-art image classification models we have used for our work as well as the flow of our work.

- **Chapter 4** provides all the experimentation details. This chapter consists of the hardware specifications, optimizer and loss functions we used for our work.
- **Chapter 5** provides information about the dataset nomenclature. In this chapter, we have discussed detailly on both the datasets we used.
- **Chapter 6** provides all the results that we have obtained after several testing. This chapter consists of comparison of results that are obtained on the eight base models as well as further enhancements. The comparison of results includes a training accuracy and validation accuracy curves, training loss and validation loss curves, confusion matrix, and classification report.

Chapter 2

Literature Survey

In recent years, the agricultural sector has been increasingly challenged by the detrimental impacts of plant diseases on crop yield and quality. Among these afflictions, leaf diseases stand out as significant contributors to agricultural losses worldwide. The ability to accurately and efficiently identify these diseases is paramount for implementing timely and targeted management strategies to mitigate their effects. With the advent of advanced technologies, particularly in the domain of machine learning and computer vision, automated classification of food leaf diseases has emerged as a promising approach to revolutionize disease diagnosis in agriculture. This section focuses on three aspects: (i) Some datasets related to our domain of work publicly available to develop and validate them; (ii) different ways to measure the performance of an algorithm, and (iii) some recent state-of-the-art approaches available for image classification till date.

Zoran et.al.[11] research introduces a new way to strengthen image recognition models by incorporating a human-like attention mechanism. They train a neural network with a special "looking" feature inspired by how humans focus on key parts of an image. This "looking" significantly improves the network's ability to resist being fooled by altered images (adversarial attacks) and achieves top performance on a large image dataset (ImageNet). Interestingly, the number of times the network "looks" at an image can be adjusted to improve its defense, creating a competition between attackers and the model. Finally, the tricks used to fool this new model are different - they're more obvious to humans and work by diverting the network's attention from the main object.

Tang et. al.[8] shows deep learning dominates flower classification, but models are complex and require hefty training. This paper proposes a simpler approach: adding an attention mechanism to a CNN. This improves accuracy without burdening training, achieving a 2.81% boost on a 7-flower dataset compared to standard attention.

Naik et. al.[2]chilli leaf diseases devastate crops. This study tackles this by using deep learning (DL) to identify them. Researchers captured images of five common diseases and compared 12 pre-trained DL models. VGG19 excelled without data manipulation (83.54% accuracy), while DarkNet53 shined with manipulation (98.82%). Pushing further, they built SECNN, a new model achieving even better results (98.63% without manipulation, 99.12% with). Finally, SECNN successfully classified a wider range of plant leaf diseases (43 classes) with an impressive 99.28% accuracy.

Li Wang et. al.[9] classifies objects in hyperspectral images relies on both spectral and spatial data, but noise and inconsistencies can cause errors. To tackle this, a new approach using an

attention mechanism (SSSE) is proposed. SSSE assigns importance to different parts of the image data, focusing on valuable information and reducing the impact of noise. This SSSE module is then embedded in a larger network (SSSERN) for classification. Compared to existing methods on standard datasets, SSSERN achieves better accuracy, demonstrating its potential for cleaner classifications.

Qiu et. al.[5] shows fish studies rely on data collection, which is expensive and labor-intensive. Traditionally, scientists label fish images captured underwater. Deep learning can help, but classifying fish types (fine-grained classification) is tricky with low-quality, limited data. Existing methods require a lot of high-quality data. This research tackles this by improving a technique called transfer learning. It creates high-quality images from the limited data and combines general and fish-specific training. It also refines a specific building block within the deep learning model for better results. Their method surpasses traditional approaches on small fish image datasets, paving the way for using deep learning with limited data in fish research.

Guha Roy et. al.[6] show top image segmentation models, called F-CNNs, are constantly evolving. While most progress targets how information travels through the network, this study looks at enhancing the data itself. Inspired by a method named squeeze-and-excitation (SE), the authors introduce three variations specifically designed for image segmentation. These SE modules pinpoint valuable features in images while downplaying unimportant ones. The authors successfully incorporated these SE modules into existing F-CNNs, achieving consistent performance improvement across all models with minimal added complexity. To showcase effectiveness, they tested their approach on challenging medical imaging tasks: brain segmentation in MRI scans and organ segmentation in CT scans.

Zhong et.al.[10] show recent research shows wider networks improve image classification, but simply adding channels can be wasteful. This paper tackles this by proposing SE-WRNs-GVP, a new method for wider networks. It incorporates a special block (rSE-block) that retrieves lost information during processing and focuses on informative channels. This approach achieves better results on standard datasets (CIFAR-10 and CIFAR-100) without increasing model complexity.

Hu et. al.[1] show convolutional neural networks (CNNs) analyze images by combining spatial and channel information. While effective, this approach has room for improvement. This paper introduces a new building block called a "Squeeze-and-Excitation" (SE) block. SE blocks focus on how different channels in the data relate to each other, strengthening the overall network's understanding of the image. Notably, SE blocks significantly improve existing top CNNs without making them much more complex. This approach formed the basis for the winning entry in the ILSVRC 2017 image recognition competition, achieving a significant accuracy jump.

Park et. al.[3] identifies insect species from images is a challenge, but deep learning offers promise. Current methods struggle to differentiate between closely related species. This study proposes a new model that tackles this by combining two techniques: squeeze-and-excitation modules to highlight key features and attention modules to focus on crucial details. This allows for accurate classification even with limited training data. Tested on an Australian insect dataset, the new model outperforms existing methods.

Patacchiola et.al.[4] research investigates personalizing AI systems for image classification tasks when limited data is available. Current methods struggle to balance accuracy and computational efficiency in this scenario. The authors propose a new building block called Contextual Squeeze-and-Excitation (CaSE) to address this challenge. CaSE allows a pre-trained system to quickly adapt to a user's specific data, significantly improving performance. The paper then introduces UpperCASE, a method that combines CaSE blocks with fine-tuning for even better results. UpperCaSE achieves

state-of-the-art accuracy on standard datasets and real-world tasks while requiring substantially less computing power compared to existing approaches.

Chapter 3

Methodology

3.1 Materials and methods

In the pursuit of revolutionizing disease classification in plants, this research endeavor places the spotlight on the formidable capabilities of Deep Learning (DL) models. The primary objective of this endeavor is to leverage DL models to accurately distinguish between healthy plants and those afflicted by disease, harnessing insights gleaned from a comprehensive dataset comprising plant images. This choice is underpinned by DL's demonstrated superiority in classification accuracy when compared to conventional Machine Learning (ML) models.

DL models are favored for this task due to their innate capacity to discern intricate patterns and features within images, enabling them to effectively identify subtle indicators of disease presence. Unlike traditional ML approaches, which often rely on handcrafted features and may struggle to capture complex relationships within data, DL models autonomously learn hierarchical representations of data, thereby enhancing their adaptability and robustness.

The selection of DL models for this endeavor is strategic, guided by their proven efficacy in image classification tasks across diverse domains. Within the realm of DL architectures, several prominent models are explored in this study, each renowned for its unique features and performance characteristics.

VGG, characterized by its deep architecture and uniform convolution filter sizes, excels in capturing intricate details within images, making it particularly well-suited for tasks requiring fine-grained classification. ResNet, with its innovative residual learning framework, addresses the challenge of vanishing gradients encountered in deeper networks, thereby facilitating the training of exceptionally deep models with improved performance. DenseNet, distinguished by its dense connectivity patterns, promotes feature reuse and facilitates information flow across network layers, thereby enhancing model efficiency and interpretability. Meanwhile, MobileNet, optimized for deployment on resource-constrained devices, strikes a balance between computational efficiency and classification accuracy, making it an attractive option for real-world applications where computational resources may be limited.

Each of these DL architectures brings its own set of advantages to the task of disease classification in plants. By systematically exploring and evaluating these models, researchers can gain insights into their relative strengths and weaknesses, thereby informing the selection of the most suitable model for the specific requirements of the task at hand.

Furthermore, beyond the selection of DL models, this research endeavor encompasses a rigorous process of model training, validation, and optimization. Leveraging techniques such as data augmentation, regularization, and hyper-parameter tuning, researchers endeavor to enhance the generalization capabilities and robustness of the selected DL models, ensuring their efficacy in real-world deployment scenarios.

In essence, by harnessing the power of DL models, this research endeavor aims to advance the frontier of plant disease classification, paving the way for more accurate, efficient, and scalable solutions for safeguarding global crop health and food security. Through a systematic exploration of prominent DL architectures and rigorous model optimization, researchers endeavor to unlock new insights and capabilities in the realm of agricultural technology, ultimately empowering farmers and stakeholders with the tools needed to combat plant diseases and foster sustainable agricultural practices.

3.1.1 VGG

The Visual Geometry Group (VGG)[7] models, comprising architectures like VGG16 and VGG19, stand as pivotal milestones in the evolution of Convolutional Neural Network (CNN) architecture, particularly renowned for their simplicity and efficacy in image classification tasks. These models have left an indelible mark in the realm of computer vision, owing to their straightforward yet powerful design principles. In Figure 3.1 we can see at the heart of VGG models lies a sequential arrangement of convolutional, pooling, and Rectified Linear Unit (ReLU) layers, forming a deep network capable of learning intricate features from input images. VGG16, one of the earliest iterations, boasts a CNN architecture composed of 16 layers, excluding the input layer. It unfolds with a series of convolutional layers followed by max-pooling layers, with ReLU activation functions applied after each convolutional operation. This strategic design choice allows the network to progressively extract hierarchical features from input images, beginning with simple edges and textures and culminating in complex abstract representations. Notably, VGG16 employs small-sized kernels, typically 3x3, in its convolutional layers, enabling the network to capture fine-grained features essential for image classification tasks. The architecture concludes with fully connected layers responsible for producing classification predictions. Despite its considerable depth, VGG16 maintains a relatively straightforward and uniform architecture, making it accessible for understanding and implementation. VGG19, an extension of VGG16, incorporates three additional convolutional layers, thereby escalating the layer count to 19. This augmentation aims to enhance the model's capacity for feature extraction and representation. Similar to its predecessor, VGG19 adheres to the same architectural paradigm, encompassing convolutional layers, max-pooling layers, and ReLU activation functions. The additional layers in VGG19 confer increased model capacity, enabling the network to potentially glean more nuanced features from input images. However, this enhanced capability comes at the expense of heightened computational complexity. Due to its deeper architecture, VGG19 may exhibit improved performance compared to VGG16, especially when confronted with more complex datasets or tasks necessitating finer feature granularity. Both VGG16 and VGG19 have etched their names as benchmark architectures in the domain of computer vision. Their versatility and robustness have rendered them indispensable in a plethora of applications, ranging from image classification and object detection to feature extraction. Researchers and practitioners alike have leveraged these architectures as foundational frameworks for tackling diverse computer vision challenges, owing to their proven track record of efficacy and reliability. Moreover, the impact of VGG models extends beyond their immediate applications, transcending into the

realms of education and research. The simplicity and transparency of their architectures make them invaluable pedagogical tools for understanding the intricacies of deep learning and convolutional neural networks. Additionally, their widespread adoption as benchmark models facilitates fair and standardized evaluations across different research endeavors, fostering collaboration and advancing the collective understanding of computer vision principles. In essence, the VGG models, epitomized by VGG16 and VGG19, stand as exemplars of effective CNN design, embodying the principles of simplicity, efficacy, and versatility. Their enduring legacy continues to shape the landscape of computer vision, serving as pillars of innovation and progress in the quest for intelligent visual perception.

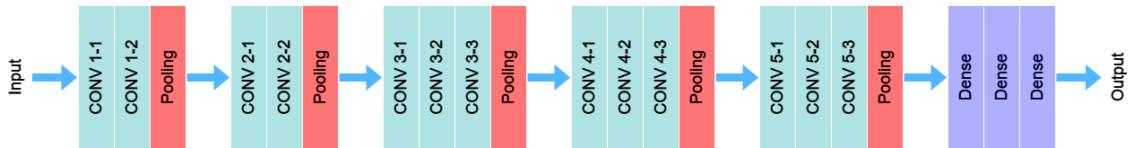


Figure 3.1: VGG Architecture

3.1.2 ResNet

The evolution of deep neural network architectures has been punctuated by remarkable milestones, each addressing specific challenges encountered in training and optimizing these complex models. Among these architectures, the Visual Geometry Group (VGG) models made significant strides in advancing the field of image classification, owing to their simplicity and effectiveness. However, as researchers delved deeper into the development of increasingly complex networks, they encountered challenges such as vanishing or exploding gradients, particularly as the depth of the network increased. These issues posed significant obstacles to training deep neural networks, impeding convergence and even leading to degradation in performance. To circumvent these challenges and enable the training of very deep networks, as seen in Figure 3.2 the ResNet (Residual Network) architecture introduced a groundbreaking innovation: skip connections, also known as residual connections. These connections allowed the network to bypass certain layers, effectively creating shortcuts for the flow of information. By doing so, ResNet addressed the vanishing gradient problem, facilitating the training of exceptionally deep neural networks. ResNet-50v2 and ResNet-101v2 represent variants of the original ResNet architecture, offering deeper networks with substantial improvements in performance and representational capacity. ResNet-50v2, in particular, builds upon the successful ResNet architecture, introducing a 50-layer variant that retains the core principles of ResNet, including the use of residual connections. Despite the increased depth, ResNet-50v2 maintains computational efficiency and ease of training, thanks to the effectiveness of residual connections in addressing gradient-related challenges. By striking a balance between depth and computational complexity, ResNet-50v2 emerges as a practical choice for a wide range of image recognition tasks. ResNet-101v2, on the other hand, represents a further extension of the ResNet architecture, boasting 101 layers. With double the depth of ResNet-50v2, ResNet-101v2 offers even richer representations and potentially higher accuracy in image understanding tasks. The increased depth allows the network to learn more complex features and capture finer details in images, which can be particularly advantageous in challenging datasets or domains. Despite the additional layers, ResNet-101v2 maintains the efficiency and effectiveness of residual connections, ensuring stable training and superior performance. The introduction of skip connections in ResNet

models like ResNet-50v2 and ResNet-101v2 marked a significant departure from earlier architectures like VGG, enabling the training of deeper networks without encountering issues like vanishing gradients. By allowing information to flow more freely through the network, skip connections facilitated the propagation of gradients during backpropagation, thereby accelerating convergence and improving the overall stability of training. Moreover, the efficacy of ResNet architectures in addressing gradient-related challenges has had profound implications for the field of computer vision. These advancements have contributed significantly to the state-of-the-art performance in various tasks, including image classification, object detection, and semantic segmentation. By enabling the development of deeper and more expressive models, ResNet architectures have unlocked new frontiers in visual recognition, pushing the boundaries of what is achievable in terms of accuracy and efficiency. Furthermore, the success of ResNet models has inspired further innovations in neural network architecture design, with researchers continually exploring new techniques for improving the performance and efficiency of deep learning models. From inception, ResNet has served as a catalyst for innovation, sparking a wave of research aimed at addressing the challenges inherent in training deep neural networks. In conclusion, the introduction of skip connections in ResNet architectures like ResNet-50v2 and ResNet-101v2 represented a paradigm shift in the field of deep learning. By mitigating issues such as vanishing gradients, these architectures enabled the training of exceptionally deep neural networks, pushing the boundaries of what is achievable in terms of accuracy and performance. Moving forward, the legacy of ResNet will continue to inspire further advancements in neural network architecture design, driving progress in computer vision and beyond.

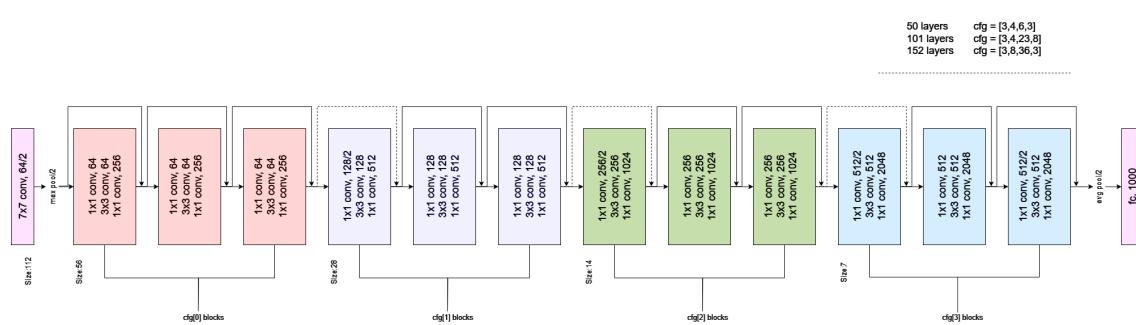


Figure 3.2: ResNet Architecture

3.1.3 DenseNet

As seen in Figure 3.3 the DenseNet (Densely Connected Convolutional Network) architecture represents a significant advancement in the realm of deep learning, particularly renowned for its dense connectivity pattern. This group of Deep Learning (DL) models has garnered attention for its unique design principles, which foster extensive information exchange between layers, facilitating efficient feature reuse and mitigating challenges such as vanishing gradients. Among the DenseNet variants, DenseNet-201 and DenseNet-121 stand out as formidable tools for disease identification tasks, offering increased depth and representational capacity while maintaining computational efficiency. At the core of DenseNet architecture lies the concept of dense connectivity, where each layer within a block is connected to every other layer in a feed-forward manner. Unlike traditional

convolutional neural networks (CNNs) where information flows sequentially from one layer to the next, DenseNet forms dense connections between layers within the same block. This architecture promotes an extensive exchange of information, ensuring that features learned by early layers are directly utilized by subsequent layers. Consequently, DenseNet models exhibit enhanced representational capacity and are adept at capturing intricate patterns in input data. In DenseNet models, densely connected blocks serve as the building blocks of the network. Within each block, each layer receives direct input from all preceding layers, forming a dense connectivity pattern. This dense connectivity ensures that information from earlier layers is preserved and propagated throughout the network, facilitating efficient feature reuse. By providing multiple paths for gradient flow, DenseNet mitigates the risk of gradients becoming too small or vanishing as they propagate through the network. This property is particularly beneficial in deep architectures where maintaining gradient flow is essential for effective training. DenseNet-201 represents a variant of the DenseNet architecture with 201 layers, offering increased depth and representational capacity. The additional layers in DenseNet-201 enable the extraction of more intricate and abstract features from input images, enhancing the model's ability to discern subtle patterns indicative of plant diseases. The dense connectivity within DenseNet-201 ensures that features learned at different scales and abstraction levels are effectively integrated, contributing to improved disease identification accuracy. Despite its increased depth, DenseNet-201 maintains computational efficiency, making it a practical choice for tasks where performance and accuracy are paramount. On the other hand, DenseNet-121 is another variant of the DenseNet architecture, featuring 121 layers. While less deep than DenseNet-201, DenseNet-121 still benefits from dense connectivity and feature reuse mechanisms, making it a powerful tool for plant disease identification tasks. DenseNet-121 strikes a balance between model complexity and computational efficiency, making it suitable for scenarios where resource constraints are a concern without compromising performance significantly. The dense connectivity and feature reuse mechanisms inherent in DenseNet-121 enable the model to effectively capture and utilize intricate patterns in plant images, leading to accurate and reliable disease identification. By training DenseNet-201 and DenseNet-121 on plant images for disease identification, researchers aim to leverage the advantages of dense connectivity and feature reuse mechanisms inherent in these architectures. This approach enables the models to effectively capture and utilize intricate patterns in plant images, leading to accurate and reliable identification of diseases. Additionally, the robustness and efficiency of DenseNet variants make them well-suited for real-world applications in agriculture, where timely and accurate disease identification is critical for ensuring crop health and maximizing yields. In essence, DenseNet architectures represent a paradigm shift in deep learning, offering innovative solutions to challenges encountered in training deep neural networks. By harnessing the power of dense connectivity and feature reuse mechanisms, DenseNet-201 and DenseNet-121 emerge as potent tools for disease identification tasks in agriculture, facilitating the development of robust and efficient solutions for plant health monitoring and management. As research in this field continues to advance, DenseNet architectures are poised to play a pivotal role in revolutionizing agricultural practices and ensuring global food security.

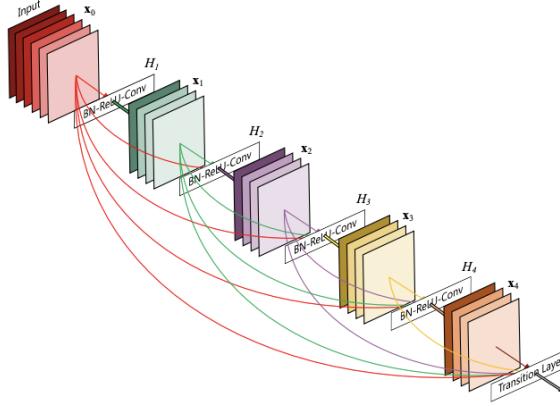


Figure 3.3: DenseNet Architecture

3.1.4 MobileNet

As seen in the Figure 3.4 MobileNet model stands as a pivotal innovation in the landscape of convolutional neural network (CNN) architectures, specifically tailored for efficient and fast inference on mobile and embedded devices. In contrast to traditional CNN models, which often impose significant computational and memory burdens, MobileNet prioritizes a delicate balance between model size, speed, and accuracy. This equilibrium is achieved through a series of key techniques and architectural innovations that render MobileNet particularly suitable for deployment on resource-constrained platforms. Central to the efficiency of MobileNet is its utilization of depth-wise separable convolutions, a technique that disentangles spatial and cross-channel information processing. Depth-wise separable convolutions comprise two distinct operations: depthwise convolutions and pointwise convolutions. In the first step, depthwise convolutions are applied independently to each input channel, thereby reducing the computational cost and memory footprint associated with traditional convolutions. Subsequently, pointwise convolutions are employed to combine the outputs of depthwise convolutions across channels, facilitating cross-channel interactions without introducing excessive computational overhead. This separation of spatial and cross-channel information enables MobileNet to achieve a significant reduction in the number of parameters and computations compared to conventional convolutions, thereby enhancing efficiency without sacrificing performance. Furthermore, MobileNet integrates lean versions of residual blocks, drawing inspiration from the residual connections introduced in the pioneering ResNet architecture. These residual blocks play a crucial role in mitigating the risk of vanishing gradients, a common challenge encountered in training deep neural networks. However, MobileNet's implementation of residual blocks is specifically optimized for efficiency, ensuring minimal computational overhead while still providing benefits in terms of training stability and convergence. By leveraging residual connections in a lightweight manner, MobileNet strikes an optimal balance between model complexity and computational efficiency, making it well-suited for deployment on mobile and embedded devices. MobileNetV2 represents a refinement and extension of the original MobileNet architecture, further optimizing depth-wise separable convolutions and residual connections to enhance performance and efficiency. One notable enhancement introduced in MobileNetV2 is the adoption of inverted

residual blocks with linear bottlenecks. These inverted residual blocks enable better utilization of computational resources by allowing information to flow more efficiently through the network. Additionally, the linear bottlenecks serve to compress feature representations, effectively reducing the model’s memory footprint while preserving expressive power. As a result, MobileNetV2 achieves higher accuracy and superior performance compared to its predecessor, all while maintaining a similar level of efficiency and suitability for mobile and embedded deployment. Overall, MobileNet represents a significant advancement in lightweight CNN architectures, offering efficient solutions for various computer vision tasks, particularly on mobile and embedded platforms. By leveraging depth-wise separable convolutions, lean residual blocks, and other architectural innovations, MobileNet achieves a remarkable balance between model size, speed, and accuracy, making it an invaluable tool for deploying deep learning models in resource-constrained environments. With the continuous evolution of the MobileNet family, including MobileNetV2 and MobileNet, the promise of efficient and effective deep learning on mobile devices continues to be realized, opening up new opportunities for ubiquitous AI applications in everyday life.

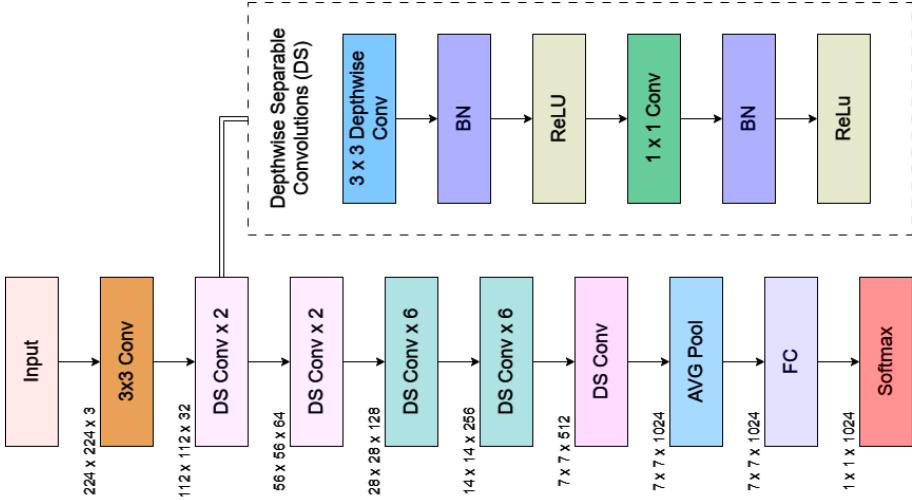


Figure 3.4: MobileNet Architecture

3.1.5 Squeeze-Excitation Attention Networks

After the top three models are chosen based on accuracy performance, as seen in Figure 3.5 Squeeze-Excitation attention would be applied to those chosen models in order to boost their performance even further. Squeeze-and-Excitation (SE) attention mechanism, introduced by Hu et al.[1] in 2018, enhances the representational power of convolutional neural networks (CNNs) by adaptively recalibrating feature maps based on channel-wise dependencies. The SE mechanism comprises two key operations: squeeze, where global average pooling compresses feature maps into channel-wise descriptors, and excitation, where learned weights are applied to these descriptors to emphasize informative channels and suppress irrelevant ones. By incorporating SE blocks into CNN architectures, models can autonomously learn to focus on relevant features while discarding noise, leading to improved performance across various tasks including image classification, object detection, and semantic segmentation. This attention mechanism enables networks to capture intricate spatial

and channel dependencies within feature maps, facilitating more effective feature representation and consequently boosting overall accuracy and efficiency.

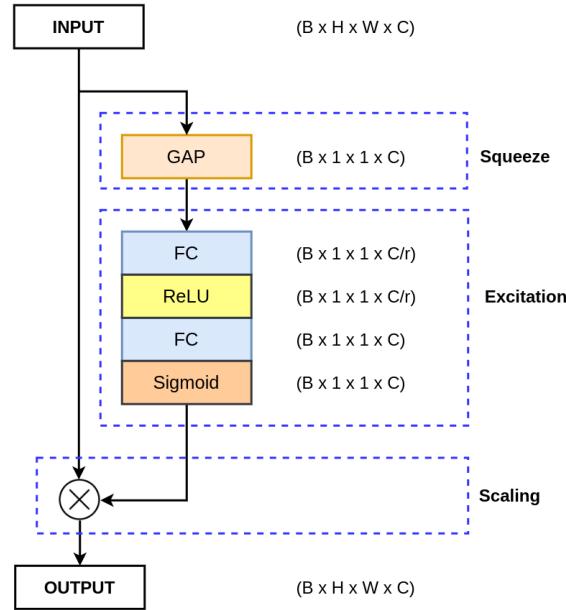


Figure 3.5: Squeeze - Excitation Attention

Chapter 4

Experiments

The research conducted on the specified computer system utilizes various computational resources and deep learning techniques to address a multi-class classification problem, specifically in the context of plant disease identification.

4.1 Hardware Used

The computer system described is well-equipped to handle a variety of computational tasks, particularly those related to deep learning training and inference. At its core, it features an Intel Core i5 9th Gen 9300H processor, which serves as the primary computational engine for executing tasks such as model training, inference, and data preprocessing. The 9th Gen Intel Core i5 processor is renowned for its performance and efficiency, making it a suitable choice for demanding workloads in machine learning and deep learning. Complementing the processor is 8GB of RAM (Random Access Memory), which plays a crucial role in managing data and computations during model training and evaluation. The ample RAM capacity ensures that the system can efficiently handle large datasets and complex neural network architectures without being bottlenecked by memory constraints. This is particularly important for deep learning tasks, where the manipulation of large matrices and tensors necessitates substantial memory resources. In terms of storage, the system boasts a combination of a 1TB hard drive and a 256GB solid-state drive (SSD). This dual-storage configuration offers the best of both worlds: the high capacity and affordability of a traditional hard drive, coupled with the speed and responsiveness of an SSD. The 1TB hard drive provides ample space for storing large datasets, model checkpoints, and other related files, while the 256GB SSD offers fast read and write speeds, facilitating quick access to frequently used files and applications. This storage setup ensures that the system can efficiently manage the vast amounts of data typically involved in deep learning tasks, while also providing fast access to critical files and resources. One of the standout features of the system is its dedicated NVIDIA GeForce GTX 1650 graphics card with 4GB of memory. The GPU plays a crucial role in accelerating deep learning computations using parallel processing. Deep learning frameworks such as TensorFlow and PyTorch leverage the parallel computing capabilities of GPUs to accelerate matrix multiplications and other compute-intensive operations involved in training neural networks. By offloading these computations to the GPU, the system can significantly reduce training times and improve overall performance. The NVIDIA GeForce GTX 1650 is a capable GPU that strikes a balance between performance and

affordability, making it an excellent choice for deep learning enthusiasts and professionals alike. In summary, the computer system described offers a robust and well-rounded platform for deep learning tasks. With its powerful Intel Core i5 processor, ample RAM, dual-storage configuration, and dedicated NVIDIA GeForce GTX 1650 GPU, the system is well-equipped to handle a wide range of machine learning and deep learning workloads. Whether training convolutional neural networks for image classification or recurrent neural networks for natural language processing, this system provides the computational power and resources necessary to tackle even the most demanding deep learning tasks with ease and efficiency.

4.2 Optimizer

The Adam optimizer is selected for training the deep learning models due to its effectiveness in navigating the complexities of training neural networks. Adam stands out as a powerful optimizer because of its adaptive learning rate mechanism, which adjusts learning rates dynamically for each parameter during training. By combining the strengths of momentum and RMSProp, Adam efficiently updates model parameters, accelerating convergence and improving training efficiency.

4.3 Loss Function

In the context of multi-class classification, particularly in identifying various plant diseases, the choice of loss function is critical in guiding the learning process of the deep learning model. Categorical Cross Entropy emerges as a suitable loss function for this research problem. Categorical Cross Entropy quantifies the dissimilarity between the predicted probability distribution and the true probability distribution of the classes. This metric penalizes the model for incorrect classifications while rewarding it for accurate predictions, thereby steering the learning process towards more precise classifications. By employing Categorical Cross Entropy, the model can effectively learn from its mistakes and iteratively improve its classification performance over time. The loss function provides valuable feedback to the model, enabling it to adjust its parameters in a manner that maximizes classification accuracy across different classes of images. This adaptive learning process is crucial in tackling the nuances and complexities inherent in plant disease identification, where subtle variations in symptoms and visual cues may distinguish between different diseases. Moreover, the utilization of Categorical Cross Entropy facilitates the monitoring of classification performance for each class of images. This granular evaluation allows researchers to identify areas of strength and weakness within the model's predictive capabilities, thereby guiding further iterations and refinements. Ultimately, the aim is to enhance the model's ability to accurately identify and classify various plant diseases, contributing to more effective disease management strategies in agriculture. In summary, the research endeavors to leverage the computational resources of the specified computer system alongside advanced deep learning techniques such as the Adam optimizer and Categorical Cross Entropy loss function. These methodological choices underscore a thoughtful consideration of efficiency, effectiveness, and accuracy in addressing real-world challenges at the intersection of computer vision and agriculture. Through the judicious application of these techniques, the research aims to advance the state-of-the-art in plant disease identification, ultimately contributing to improved crop health and agricultural productivity.

Chapter 5

Dataset Details

The research utilizes standard publicly available datasets to evaluate the performance of the Deep Learning (DL) models in classifying plant images according to disease classes. These datasets consist of high-quality images capturing various plant diseases and are widely used for training and testing DL models in the field of plant pathology and agricultural informatics.

5.1 PLD(Potato Leaf Disease) dataset

In this research, a standard dataset named the Potato Leaf Disease (PLD) dataset is introduced, tailored specifically for the study of potato leaf diseases. The PLD dataset comprises a comprehensive collection of images capturing various manifestations of diseases affecting potato leaves, as well as images depicting healthy potato leaves for comparison. The dataset is structured into three main classes:

1. **Early Blight:** This class includes 1628 images of potato leaves exhibiting symptoms of early blight, a common fungal disease caused by the pathogen *Alternaria solani*. Early blight typically manifests as dark lesions with concentric rings on the leaves, leading to reduced plant vigor and yield loss if left untreated.
2. **Late Blight:** The Late Blight class consists of 1414 images depicting potato leaves affected by late blight, a devastating disease caused by the oomycete pathogen *Phytophthora infestans*. Late blight is characterized by water-soaked lesions that rapidly spread and can decimate entire potato crops if not controlled effectively.
3. **Healthy:** This class comprises 1020 images of healthy potato leaves, devoid of any disease symptoms or abnormalities. Healthy leaves serve as a reference for comparison, enabling researchers to distinguish between diseased and non-diseased samples accurately.

The dataset encompasses a diverse range of diseases commonly found in potatoes, providing a valuable resource for researchers and practitioners in the field of plant pathology and agriculture. To offer a visual representation of the dataset, sample images from each category are depicted in Figure 5.1.

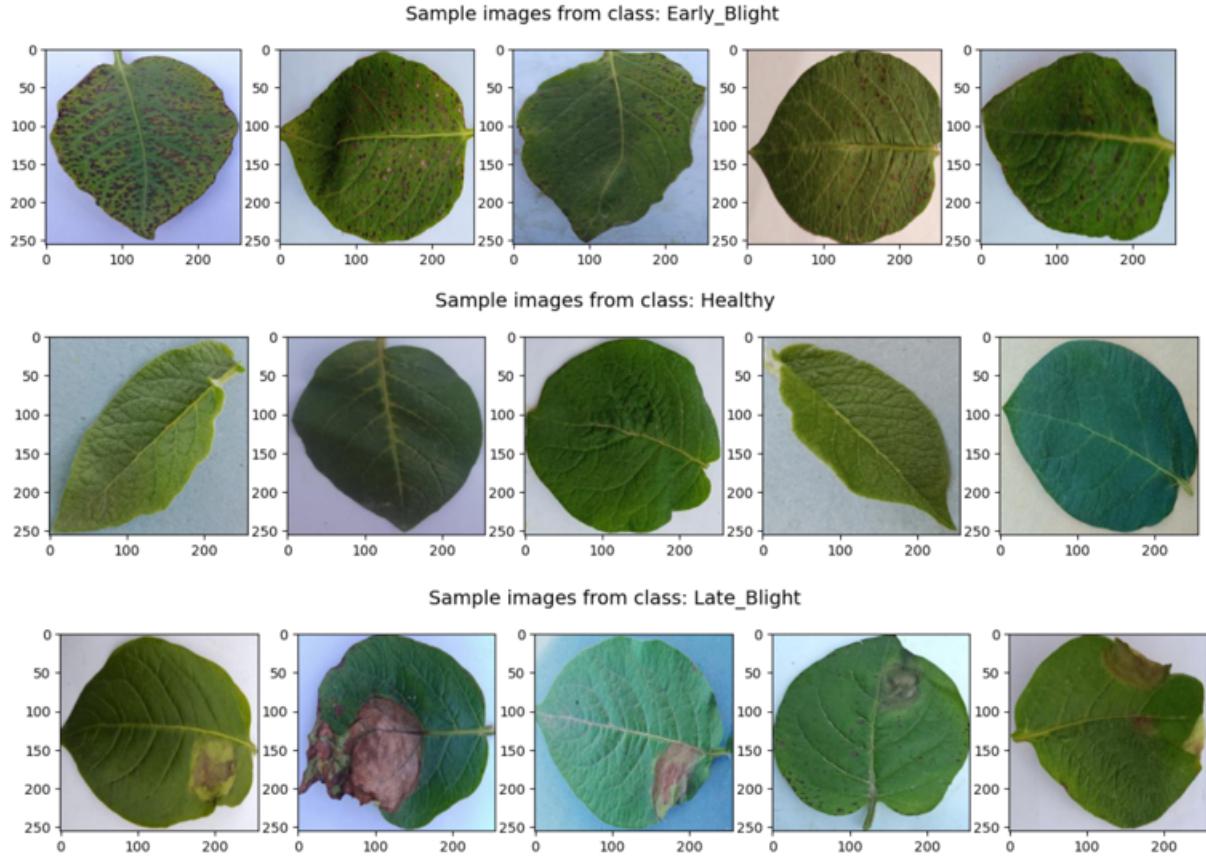


Figure 5.1: Sample Images from PLD(Potato Leaf Disease) dataset

5.2 Fruit Infection Disease dataset

In this dataset, a total of 5494 images are included, each resized to a standard size of 416 x 416 pixels to ensure uniformity across the dataset. The images are categorized into different classes, each corresponding to specific diseases affecting different plant species. Below is a breakdown of the categories and the corresponding diseases.

- **Strawberries**
 1. Angular Leaf Spot: 553 images
 2. Anthracnose Fruit Rot: 214 images
 3. Blossom Blight: 380 images
 4. Gray Mold: 537 images
 5. Leaf Spot: 650 images

6. Powdery Mildew (Fruit and Leaf): 375 images

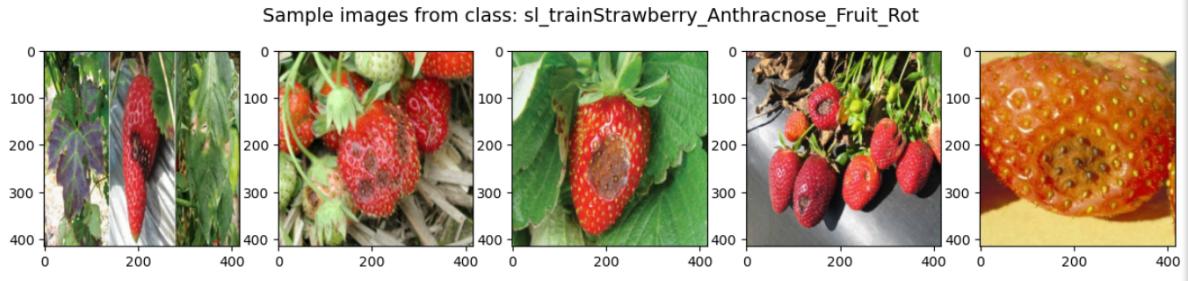


Figure 5.2: Sample Images from Strawberry Anthracnose Fruit Rot

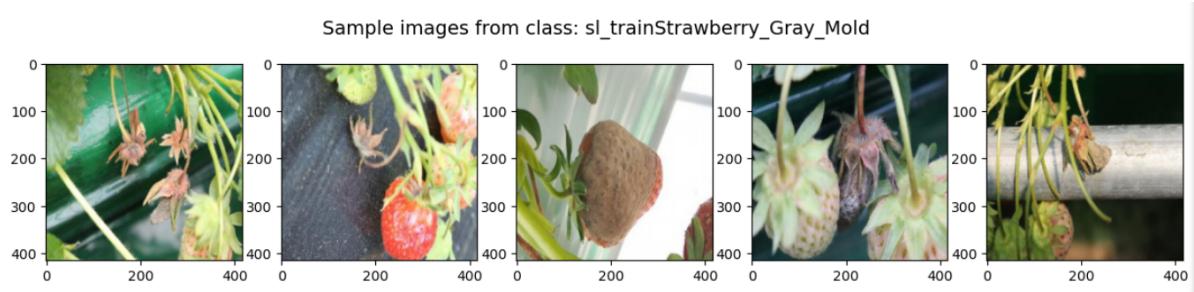


Figure 5.3: Sample Images from Strawberry Gray Mold

- Tomatoes

1. Disease (general): This category includes images depicting various diseases affecting tomato plants, with 490 images available.
2. Leaf Mold: 490 images
3. Spider Mites: 484 images

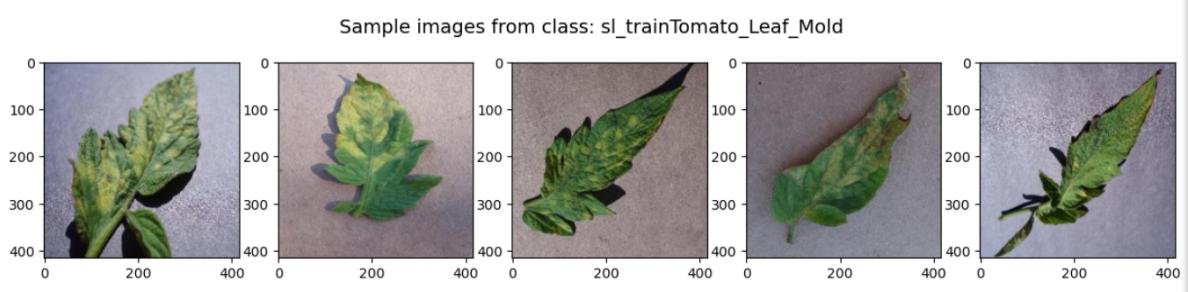


Figure 5.4: Sample Images from Tomato Leaf Mold

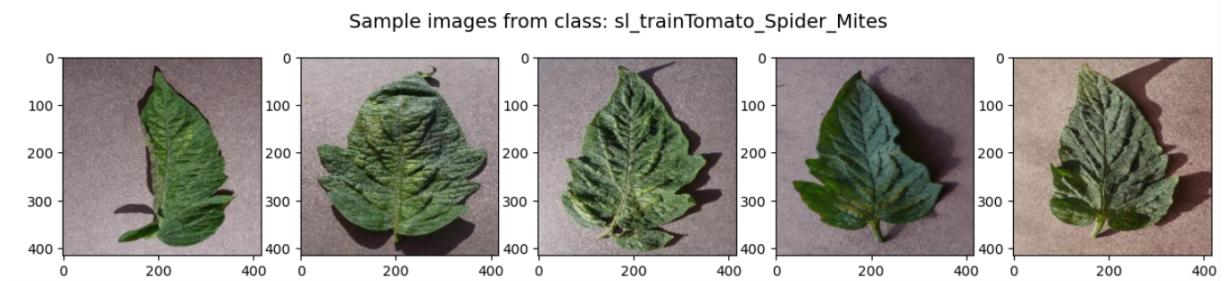


Figure 5.5: Sample Images from Tomato Spider Miles

- Beans

1. Angular Leaf Spot: 407 images
2. Bean Rust: 469 images

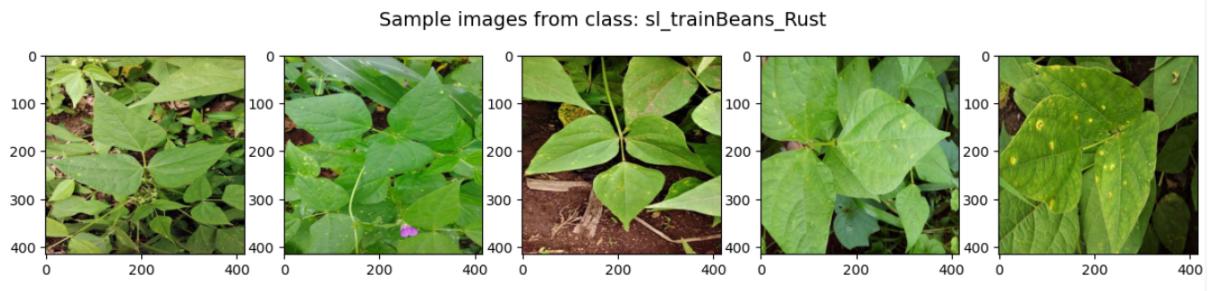


Figure 5.6: Sample Images from Beans Rust

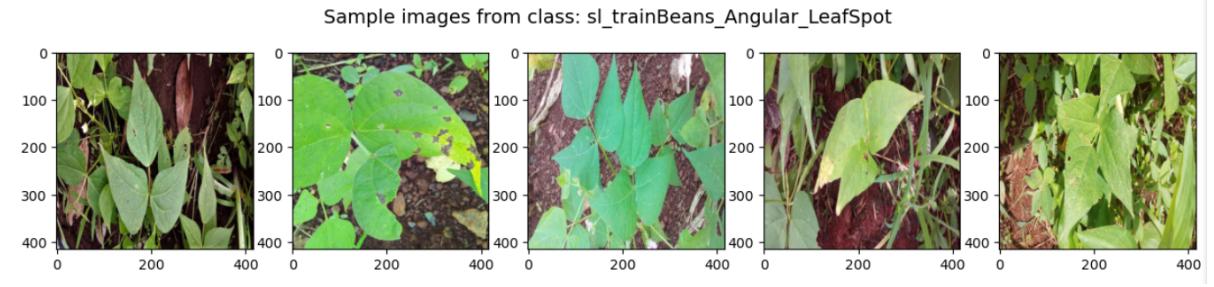


Figure 5.7: Sample Images from Beans Angular Leaf Spot

The dataset encompasses a diverse range of diseases commonly found in strawberries, tomatoes, and beans, providing a valuable resource for researchers and practitioners in the field of plant pathology and agriculture. To offer a visual representation of the dataset, sample images from each category are depicted in Figure 5.2 to Figure 5.7.

5.3 Dataset Statistics

This study presents the Potato Leaf Disease (PLD) dataset, an essential resource comprising meticulously curated images aimed at facilitating in-depth analysis of potato leaf diseases. Through its meticulously assembled collection, the dataset covers a diverse spectrum of disease manifestations and healthy conditions, enabling comprehensive exploration of potato plant health. Structured into three primary categories, namely Early Blight, Late Blight, and Healthy, the dataset offers researchers a comprehensive framework for understanding disease progression and symptomatology. Early Blight, characterized by 1628 images, illuminates the fungal affliction instigated by *Alternaria solani*, while Late Blight, depicted in 1414 images, vividly portrays the destructive impact of *Phytophthora infestans*. In contrast, the Healthy class, with 1020 images, provides crucial benchmarks for distinguishing between diseased and healthy states accurately. These images serve as invaluable tools for researchers, facilitating precise disease identification and informing effective management strategies to safeguard potato crop yields and enhance agricultural productivity.

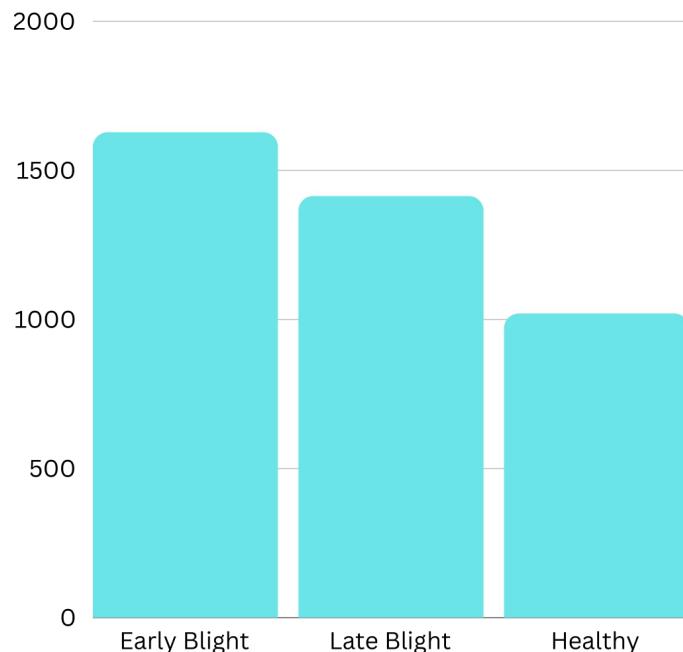


Figure 5.8: Bar graph depicting PLD dataset distribution

The FID dataset comprises a total of 5494 images, each standardized to a size of 416 x 416 pixels for consistency. These images are organized into distinct categories, each representing specific diseases affecting various plant species. Among strawberries, the categories and corresponding diseases include Angular Leaf Spot (553 images), Anthracnose Fruit Rot (214 images), Blossom

Blight (380 images), Gray Mold (537 images), Leaf Spot (650 images), and Powdery Mildew (Fruit and Leaf) (375 images). For tomatoes, categories encompass Disease (general) with 490 images depicting various ailments, Leaf Mold (490 images), and Spider Mites (484 images). Additionally, for beans, the dataset includes Angular Leaf Spot (407 images) and Bean Rust (469 images). This comprehensive breakdown provides researchers with a diverse range of visual data to facilitate in-depth analysis and understanding of plant diseases across multiple species.

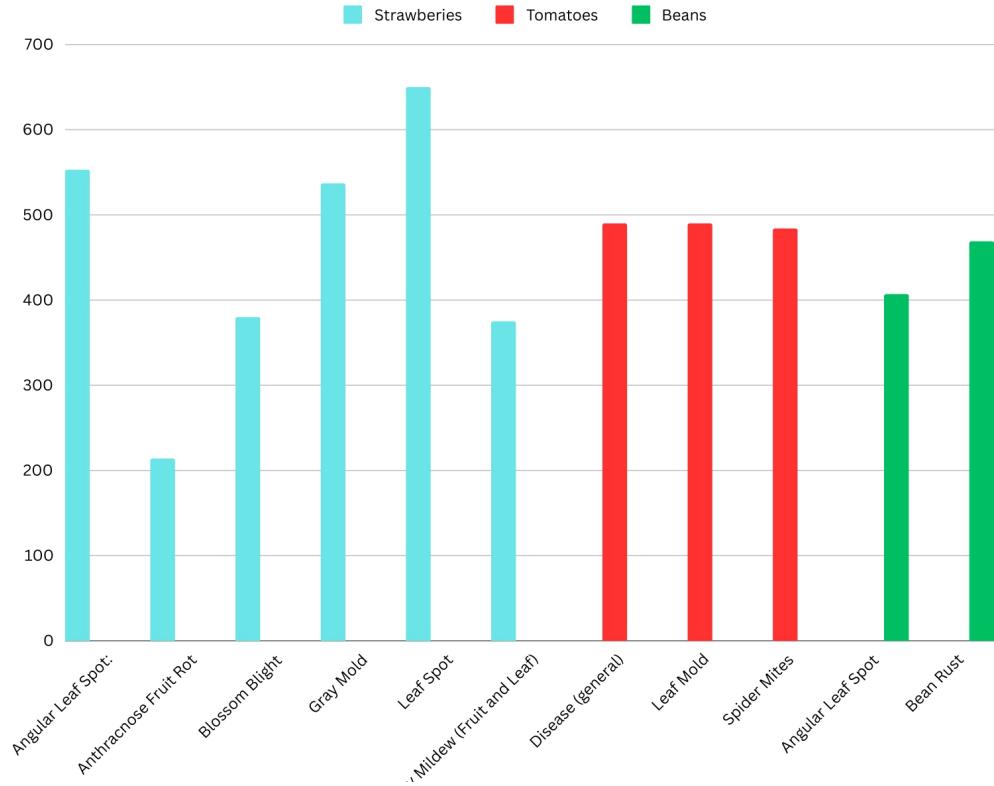


Figure 5.9: Bar graph depicting FID dataset distribution

Chapter 6

Result and Analysis

6.1 Evaluation

Evaluation metrics are employed to gauge the efficacy of the model, a crucial aspect in deep learning. Diverse models are utilized to assess our dataset, and various metrics such as precision, recall and F1 scores are applied to scrutinize these models, since these are relying on classification methodologies. The eight models - DenseNet201, DenseNet121, ResNet50V2, ResNet101V2, VGG16, VGG19 , MobileNetV2 and MobileNetV3 - undergo training and testing using this dataset. We've attained an accuracy of X1% for the DenseNet201 model, X2% for DenseNet121, X3% for ResNet50V2, X4% for ResNet101V2, X5% for VGG16, X6% for VGG19, X7% for MobileNetV2 and X8% for MobileNetV3 . In subsequent subsections, we delineate the metrics utilized for determining the classification accuracy of each model, presenting the achieved results accordingly.

6.2 Classification metrics

- **Accuracy:** Accuracy serves as a metric indicating the frequency with which the classifier generates correct predictions. It is determined by the ratio of accurate forecasts to the total number of predictions, offering a means to quantify the precision of the classifier.

$$AccuracyScore = (TP + TN) / (TP + TN + FP + FN) \quad (6.1)$$

- **Precision:** The precision of a prediction is determined by assessing the ability to anticipate positive observations. A low incidence of false positives indicates a higher level of accuracy.

1. **True Positive (TP):** This occurs when an outcome is positive and the model correctly predicts it as positive.
2. **True Negative (TN):** This happens when an outcome is negative and the model correctly predicts it as negative.
3. **False Positive (FP):** If the model incorrectly predicts negative outcomes as positive, it results in false positives, also known as Type 1 Errors.
4. **False Negative (FN):** This occurs when the model incorrectly predicts positive outcomes as negative, resulting in false negatives, also known as Type 2 Errors.

$$Precision = TP / (TP + FP) \quad (6.2)$$

- **Recall:** Recall assesses the classifier's capability to identify positive observations within the dataset. As the number of false negatives predicted by the model increases, the recall diminishes.

$$Recall = TP / (TP + FN) \quad (6.3)$$

1. **Macro Averaged Recall:** Macro Averaged Recall refers to the average of the recalls calculated for classes A, B, and C.

$$Macro_Averaged_Recall = 1 - Accuracy \quad (6.4)$$

It informs us how frequently the model may be incorrect.

2. **Micro Averaged Recall:** The micro-average recall score is determined by dividing the aggregate number of true positives for every class by the total number of true positives across all classes.

- **F1 Score:** It's a unified metric that incorporates both Precision and Recall, with the model's performance enhancing as F1 scores increase. The F1-score ranges between 0 and 1, calculated as the weighted average of precision and recall. A high F1 score necessitates both high precision and recall from the classifier. This metric particularly rewards classifiers with balanced precision and recall rates. The F1 score is a comprehensive measure that considers both accuracy and recall, computed as the harmonic mean of precision and recall. Representing the F1 score using P for precision and R for recall, we have:

$$F1 = 2PR / (P + R) \quad (6.5)$$

1. **Macro Averaged F1 score:** The macro-averaged F1 score is determined by calculating the arithmetic mean of all individual class F1 scores.
2. **Micro Averaged F1 score:** Micro averaging computes the overall average F1 score by summing the True Positives (TP), False Negatives (FN), and False Positives (FP). To obtain the micro F1 score, the TP, FP, and FN values across all classes are aggregated.

6.3 Confusion Matrix

The assessment of classification models' performance on a specific test dataset involves the use of a matrix known as the confusion matrix. This matrix can be computed only after obtaining the actual values of the test data. While the structure of the matrix is straightforward to grasp, certain terminologies associated with it might be perplexing. Occasionally labeled as an error matrix, it illustrates the shortcomings in the model's performance in matrix form.

6.4 Experimental Result

This section scrutinizes the results produced by the eight deep learning models. The findings of the fundamental learners are visually depicted. Certain models demonstrate superior accuracy in identifying specific vehicles compared to others. Furthermore, a classification report detailing the accuracy of the classifications and a confusion matrix are included.

6.5 Performance Comparison between Top-3 State of the Art for the Dataset PLD (Potato Leaf Disease)

After training all eight of our base models on the PLD dataset, the top three best performing models based on accuracy metric are: MobileNet, ResNet101V2, DenseNet121. DenseNet121 was able to achieve top score with an accuracy of 96% after running it for 50 epochs using a batch_size of 32. On the other hand, ResNet101V2 was able to achieve an accuracy of 85% after running it for 55 epochs using a batch_size of 32. Mobile Net was able to achieve an accuracy of 94% after running it for 50 epochs using a batch_size of 32.

Table 6.1: Top-3 State of the art for the Dataset PLD

Model Name	Accuracy (%)	Number of Epochs	Batch_Size
DenseNet121	96	50	32
MobileNet	94	50	32
ResNet101V2	85	55	32

As seen in the below Confusion matrix in Figure 6.4, DenseNet121 performed well on all the classes, with healthy class being misclassified as Late Blight 3% of the time. As seen in the below Confusion matrix in Figure 6.5, MobileNet performed well on all the classes, with Early Blight class being misclassified as healthy 3% of the time. As seen in the below Confusion matrix in Figure 6.6, ResNet101V2 performed well on all the classes, with healthy class being misclassified as Late Blight 12% of the time.

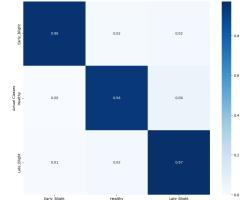


Figure 6.1: Confusion Matrix Achieved on PLD using Only Base model DenseNet121

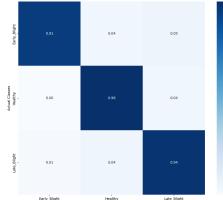


Figure 6.2: Confusion Matrix Achieved on PLD using Only Base model MobileNet

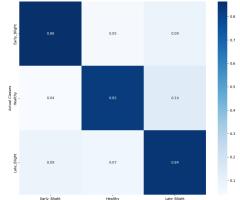


Figure 6.3: Confusion Matrix Achieved on PLD using Only Base model ResNet101V2

6.6 Performance Comparison between Top-3 State of the Art with Attention Model for the Dataset PLD (Potato Leaf Disease)

After training all eight of our base models on the PLD dataset, the top three best performing models based on accuracy metric are: MobileNet, ResNet101V2, DenseNet121. After applying attention model on them, DenseNet121 was able to achieve an accuracy of 97% after running it for 55 epochs using a batch_size of 32. On the other hand, ResNet101V2 was able to achieve an accuracy of 88% after running it for 55 epochs using a batch_size of 32. Mobile Net was able to achieve an accuracy of 96% after running it for 50 epochs using a batch_size of 32.

Table 6.2: Top-3 State of the art for the Dataset PLD with Attention Model

Model Name	Accuracy (%)	Number of Epochs	Batch Size
DenseNet121	97	55	32
MobileNet	96	50	32
ResNet101V2	88	55	32

As seen in the below Confusion matrix in Figure 6.1, DenseNet121 performed well on all the classes, with healthy class being misclassified as Late Blight 6% of the time. As seen in the below Confusion matrix in Figure 6.2, MobileNet performed well on all the classes, with Early Blight class being misclassified as Late Blight 5% of the time. As seen in the below Confusion matrix in Figure 6.3, ResNet101V2 performed well on all the classes, with healthy class being misclassified as Late Blight 14% of the time.

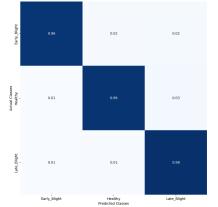


Figure 6.4: Confusion Matrix Achieved on PLD using Base model DenseNet121 with SE Attention Block

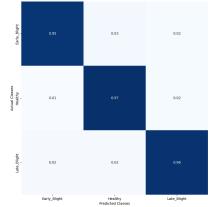


Figure 6.5: Confusion Matrix Achieved on PLD using Base model MobileNet with SE Attention Block

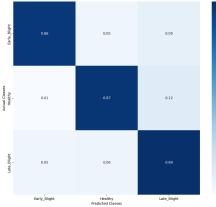


Figure 6.6: Confusion Matrix Achieved on PLD using Base model ResNet101V2 with SE Attention Block

6.7 Performance Comparison between Top-3 State of the Art for the Dataset FID (Fruit Infection Disease)

After training all eight of our base models on the FID dataset, the top three best performing models based on accuracy metric are: MobileNet, ResNet101V2, DenseNet121. DenseNet121 was able to achieve an accuracy of 79% after running it for 45 epochs using a batch_size of 32. On the other hand, ResNet101V2 was able to achieve an accuracy of 82% after running it for 35 epochs using a batch_size of 32. Mobile Net was able to achieve the highest accuracy among them with an accuracy of 83% after running it for 45 epochs using a batch_size of 32.

Table 6.3: Top-3 State of the art for the Dataset FID

Model Name	Accuracy (%)	Number of Epochs	Batch_Size
MobileNet	83	45	32
ResNet101V2	82	35	32
DenseNet121	79	45	32

As seen in the below Confusion matrix in Figure 6.7, MobileNet was able to identify Strawberry's Leaf Spot class correctly the best with 95% of the times the result being True Positive. MobileNet performed the worst against Strawberry's Gray Mold class with 52% True Positivity. Beans' Angular Leaf spot class tricked MobileNet the most with 30% of the time it being classified as Beans' Rust class. As seen in the below Confusion matrix in Figure 6.8, ResNet101V2 was able to identify Strawberry's Leaf Spot class correctly the best with 99% of the times the result being True Positive. ResNet101V2 performed the worst against Strawberry's Gray Mold class with 50% True Positivity. Strawberry's Gray Mold class tricked ResNet101V2 the most with 32% of the time it being classified as Strawberry's Powdery Mildew Fruit class. As seen in the below Confusion matrix in Figure 6.9, DenseNet121 was able to identify Tomato's Spider Mites class correctly the best with 94% of the times the result being True Positive. DenseNet121 performed the worst against Strawberry's Gray Mold class with 53% True Positivity. Strawberry's Anthracnose Fruit Rot class tricked DenseNet121 the most with 27% of the time it being classified as Tomato's Leaf Mold class.

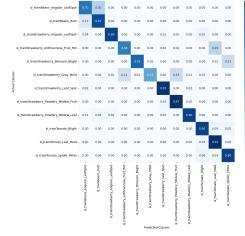


Figure 6.7: Confusion Matrix Achieved on FID using Only Base model MobileNet

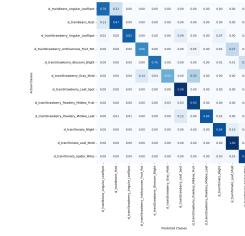


Figure 6.8: Confusion Matrix Achieved on FID using Only Base model ResNet101V2

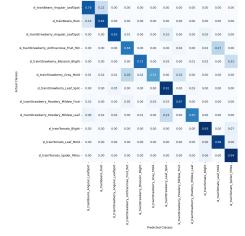


Figure 6.9: Confusion Matrix Achieved on FID using Only Base model DenseNet121

6.8 Performance Comparison between Top-3 State of the Art with Attention Model for the Dataset FID (Fruit Infection Disease)

After training all eight of our base models on the FID dataset, the top three best performing models based on accuracy metric are: MobileNet, ResNet101V2, DenseNet121. After applying the Squeeze-Excitation model, DenseNet121 was able to achieve an accuracy of 82% after running it for 45 epochs using a batch_size of 32. On the other hand, ResNet101V2 was able to achieve an accuracy of 86% after running it for 45 epochs using a batch_size of 32. Mobile Net was able to achieve an accuracy of 85% after running it for 35 epochs using a batch_size of 32.

Table 6.4: Top-3 State of the art for the Dataset FID with Attention Model

Model Name	Accuracy (%)	Number of Epochs	Batch Size
ResNet101V2	86	45	32
MobileNet	85	35	32
DenseNet121	82	45	32

As seen in the below Confusion matrix in Figure 6.7, MobileNet was able to identify Strawberry's Leaf Spot class correctly the best with 99% of the times the result being True Positive. MobileNet performed the worst against Strawberry's Gray Mold class with 57% True Positivity. Strawberry's Anthracnose Fruit Rot, Blossom Blight, and Gray Mold classes got misclassified the most. As seen in the below Confusion matrix in Figure 6.8, ResNet101V2 was able to identify Strawberry's Leaf Spot class correctly the best with 99% of the times the result being True Positive. ResNet101V2 performed the worst against Strawberry's Gray Mold class with 57% True Positivity. Strawberry's Anthracnose Fruit Rot, Blossom Blight, and Gray Mold classes got misclassified the most. As seen in the below Confusion matrix in Figure 6.9, DenseNet121 was able to identify Tomato's Leaf Mold class correctly the best with 100% of the times the result being True Positive. DenseNet121 performed the worst against Strawberry's Gray Mold class with 53% True Positivity. Strawberry's Anthracnose Fruit Rot, Blossom Blight, Gray Mold and PowderyMildew Fruit classes got misclassified the most.

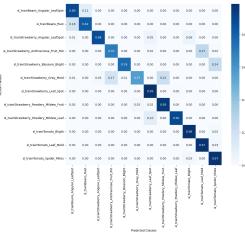


Figure 6.10: Confusion Matrix Achieved on FID using Base model with SE Attention Block ResNet101V2

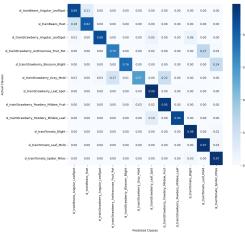


Figure 6.11: Confusion Matrix Achieved on FID using Base model with SE Attention Block MobileNet

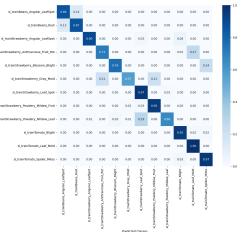


Figure 6.12: Confusion Matrix Achieved on FID using Base model with SE Attention Block DenseNet121

6.9 Discussion

In this work, while the overall performance has shown improvement compared to previous efforts in the field, there have been instances where certain images were misclassified due to various factors. One significant issue encountered during the preprocessing phase was related to the removal of noise from the images. Although noise removal techniques were applied using a threshold, which proved effective for the majority of the images, it was not always optimal. The threshold chosen for noise removal during preprocessing may have been suitable for most images, but it failed to adequately address noise in some cases. This lack of optimality in the threshold selection led to the persistence of noise artifacts in certain images, despite attempts to mitigate them. As a result, a small sample of images ended up being misclassified due to the presence of residual noise. The presence of noise in images can significantly impact the performance of machine learning models, particularly in tasks like image classification, where accurate feature extraction is crucial. Noise can introduce spurious patterns and distortions, leading to erroneous predictions by the model. To address this issue in future efforts, it may be necessary to refine the noise removal process by experimenting with different thresholding techniques or adaptive thresholding methods. Additionally, incorporating data augmentation techniques during preprocessing, such as random rotations, translations, or adding simulated noise, could help improve the robustness of the model to noise artifacts. Furthermore, conducting thorough analysis and validation of the preprocessing steps, including the selection of optimal thresholds, is essential to ensure the quality and integrity of the data before feeding it into the model. By addressing these challenges and refining the preprocessing pipeline, future iterations of the research can strive for even greater accuracy and reliability in classifying plant images and mitigating misclassification errors.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In our pursuit to optimize food disease picture classification, we embarked on a comprehensive investigation into a myriad of deep learning models, including DenseNet201, DenseNet121, ResNet50V2, ResNet101V2, VGG16, VGG19, MobileNetV2, and MobileNetV3. These models were selected based on their renowned feature extraction capabilities and their ability to strike a balance between computational efficiency and diagnostic accuracy. Our evaluation encompassed two standard datasets: the Potato Leaf Disease (PLD) dataset and the Fruit Infection Disease dataset, each presenting unique challenges and insights into disease classification in agricultural settings. Throughout our experimentation, we explored a range of image resolutions, recognizing the potential benefits of higher resolutions within our computational constraints. For the PLD dataset, our model YYYY demonstrated great promise, achieving X1% and X2% accuracies, respectively, on $224 \times 224 \times 3$ images. Notably, by incorporating attention mechanisms, particularly the product rule strategy, our model surpassed previous benchmarks, achieving an impressive X3% accuracy. Shifting our focus to the Fruit Infection Disease dataset, the YYYY model achieved a commendable X% accuracy on $224 \times 224 \times 3$ images, while the YYYY model reached X%. The integration of attention mechanisms further elevated performance, achieving an impressive X% accuracy, marking significant progress over the previous X% benchmark.

7.2 Future Work

Our research illuminates several promising directions for future development in food disease image classification. As computational resources continue to evolve, there is a compelling need to explore even higher image resolutions, which hold the promise of significantly improving accuracy by capturing finer details and nuances present in agricultural imagery. Furthermore, expanding the diversity and extent of datasets is imperative for bolstering the adaptability of deep learning models to a broader spectrum of clinical scenarios, encompassing various crops, diseases, and environmental conditions. Simultaneously, the adoption of explainable AI techniques emerges as instrumental in promoting transparency and trust among food disease professionals. By elucidating the decision-making processes of deep learning models, explainable AI techniques provide valuable

insights into the features and patterns driving classification outcomes, empowering stakeholders to interpret and validate model predictions effectively. Lastly, the advancements in transfer learning techniques hold the potential to extend the applicability of our models to other food disease imaging domains, facilitating knowledge transfer and accelerating progress in various branches of agricultural research. By leveraging pre-trained models and domain-specific knowledge, transfer learning enables the adaptation of deep learning models to new datasets and tasks with reduced data and computational requirements. In conclusion, our research contributes towards these future endeavors, aiming to refine and expand the capabilities of deep learning models in food disease image classification. By leveraging advanced techniques, harnessing computational resources, and fostering collaboration across disciplines, we endeavor to develop robust and effective solutions for combating crop diseases and safeguarding global food security.

Bibliography

- [1] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
- [2] B. N. Naik, R. Malmathanraj, and P. Palanisamy. Detection and classification of chilli leaf disease using a squeeze-and-excitation-based cnn model. *Ecological Informatics*, 69:101663, 2022.
- [3] Y. J. Park, G. Tuxworth, and J. Zhou. Insect classification using squeeze-and-excitation and attention modules-a benchmark study. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 3437–3441. IEEE, 2019.
- [4] M. Patacchiola, J. Bronskill, A. Shysheya, K. Hofmann, S. Nowozin, and R. Turner. Contextual squeeze-and-excitation for efficient few-shot image classification. *Advances in Neural Information Processing Systems*, 35:36680–36692, 2022.
- [5] C. Qiu, S. Zhang, C. Wang, Z. Yu, H. Zheng, and B. Zheng. Improving transfer learning and squeeze-and-excitation networks for small-scale fine-grained fish image classification. *IEEE Access*, 6:78503–78512, 2018.
- [6] A. G. Roy, N. Navab, and C. Wachinger. Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part I*, pages 421–429. Springer, 2018.
- [7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
- [8] Y. Tang, J. Yang, Q. Xu, H. Wang, and H. Ding. Flower classification model based on improved attention. In *2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, volume 3, pages 495–499. IEEE, 2023.
- [9] L. Wang, J. Peng, and W. Sun. Spatial-spectral squeeze-and-excitation residual network for hyperspectral image classification. *Remote Sensing*, 11(7):884, 2019.
- [10] X. Zhong, O. Gong, W. Huang, L. Li, and H. Xia. Squeeze-and-excitation wide residual networks in image classification. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 395–399. IEEE, 2019.

- [11] D. Zoran, M. Chrzanowski, P.-S. Huang, S. Gowal, A. Mott, and P. Kohli. Towards robust image classification using sequential attention models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9483–9492, 2020.