Apple Leaf Diseases Detection and Classification using Deep Learning

By

Wasi Ullah

00000327455

Supervisor

Dr. Kashif Javed

Department of Robotics and Artificial Intelligence
School of Mechanical and Manufacturing Engineering (SMME)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan

14 September 2023

Apple Leaf Diseases Detection and Classification using Deep Learning

By

Wasi Ullah

00000327455

Supervisor

Dr. Kashif Javed

Co-supervisor

Dr. Sara Babar

A thesis submitted in conformity with the requirements for
the degree of *Master of Science* in
Robotics and Intelligent Machines Engineering

Department of Robotics and Artificial Intelligence
School of Mechanical and Manufacturing Engineering (SMME)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
14 September 2023

Declaration

I declare that this thesis titled "Apple Leaf Diseases Detection and Classification using Deep Learning" is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Wasi Ullah,
Reg no. 00000327455

Plagiarism Certificate

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.

Wasi Ullah,
Reg no. 00000327455

Signature of Supervisor

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by **Regn No. 00000327455 Wasi Ullah** of **School of Mechanical & Manufacturing Engineering (SMME) (SMME)** has been vetted by undersigned, found complete in all respects as per NUST Statues/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis titled. **Apple Leaf Diseases Detection and Classification using Deep Learning**

Signature: Kashif

Name (Supervisor): Kashif Javed

Date: 05 - Sep - 2023

Signature (HOD):

Date: 05 - Sep - 2023

Signature (DEAN):

Date: 05 - Sep - 2023

Form TH-4

National University of Sciences & Technology (NUST)

MASTER'S THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: Wasi Ullah (00000327455) Titled: Apple Leaf Diseases Detection and Classification using Deep Learning be accepted in partial fulfillment of the requirements for the award of MS in Robotics & Intelligent Machine Engineering degree.

Examination Committee Members

1. Name: Sara Babar Signature:

2. Name: Khawaja Fahad Iqbal Signature:

3. Name: Yasar Ayaz Signature:

Supervisor: Kashif Javed Signature: Kashif

Date: 05 - Sep - 2023

Head of Department

05 - Sep - 2023

Date

COUNTERSINGED

05 - Sep - 2023

Date

Dean/Principal

Proposed Certificate for Plagiarism

It is certified that MS Thesis Titled __Apple Leaf Disease Detection and Classification using Deep Learning __by __Wasi Ullah_ has been examined by us. We undertake the follows:

a. Thesis has significant new work/knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
b. The work presented is original and own work of the author (i.e. there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.
c. There is no fabrication of data or results which have been compiled/analyzed.
d. There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
e. The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

Name & Signature of Supervisor

__Dr. Kashif Javed_

Signature : ________________

Copyright Notice

This thesis is dedicated to *my beloved parents*

Abstract

Timely diagnosis and identification of apple leaf diseases are imperative for preventing the spread of diseases and ensuring the sound development of the apple industry. Convolutional Neural Networks (CNNs) have achieved phenomenal success in the area of leaf disease detection, which can greatly benefit the agriculture industry. However, their large model size and intricate design continue to pose a challenge when it comes to deploying these models on lightweight devices. Although several successful models (e.g. EfficientNet and MobileNet) have been designed to adapt to resource-constrained devices, these models have not been able to achieve significant results in leaf disease detection task and leave a performance gap behind. This research gap has motivated us to develop an apple leaf disease detection model that can not only be deployed on lightweight devices but also can outperform the existing models. In this work, we propose AppVit, a hybrid vision model, combining the features of convolution and multi-head self-attention, to compete with the best performing models. Specifically, we begin by introducing the convloution blocks that narrows-down the size of the feature maps and helps the model to encode local features progressively. Then, we stack VIT blocks in combination with convolution blocks allowing the network to capture non-local dependencies and spatial patterns. Embodied with these designs and a hierarchical structure, AppVIT demonstrates excellent performance on apple leaf disease detection task. Specifically, it achieves 96.38 % accuracy on plant pathology 2021 - FGVC8 with about 1.3 million parameters, which is 11.3% and 4.3% more accurate than ResNet-50 and EfficientNet-B3. The precision, recall and f-score of our proposed model on apple leaf disease detection and classification are 0.967, 0.959, 0.963 respectively.

Keywords: *Transformers, Convolutional Neural Networks (CNNs), Leaf Disease Detection, Vision Transformer, Multi-head self-attention.*

Acknowledgments

I would like to express my deepest appreciation to all those who provided me the possibility to complete this thesis. A special gratitude I give to my academic advisor, Dr. Kashif Javed, whose expertise, understanding, and patience added considerably to my research experience. I appreciate their constructive comments and suggestions during the development phase of this research.

I'm heartily thankful to my committee members who provided insight and expertise that greatly assisted the research.

Furthermore, I would also like to acknowledge with much appreciation the crucial role of the staff members of SMME, who gave the permission to use all required equipment and the necessary materials to complete the task. A big thank you to my colleagues who helped me gather data and gave their valuable suggestions.

I'm also grateful to my family and friends for their unconditional support, both emotionally and financially. Their encouragement when the times got tough was invaluable.

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

CNN	Convolutional Neural Networks
VIT	Vision Transformer
MHSA	Multi-head self-attention
ALDD	Apple leaf disease detection dataset

CHAPTER 1

Introduction

1.1 Introduction

1.1.1 Background

Apples are one of the most widely cultivated and consumed fruits worldwide. Apples because of their high nutritive and remedial value are considered to be the most productive fruit. The antioxidant effects, due to the presence of rich amount fiber and phytochemicals, apple helps guard a cell's DNA against oxidative harm, which can lead to cancer [1]. Moreover, these chemicals present in the apple prevents the growth of new cancer cells and inhibit the expansion of existing ones [1]. Apple is rich in vitamin C, sodium, potassium, fiber, phosphorus, calcium, and iron. [2] have demonstrated that eating more apples can help lower the risk of stroke. Besides their significant nutritive benefits, apples also play a vital role in the economies of agrarian countries. In Pakistan, apple is the fourth largest fruit crop in terms of production. In 2020, apple production in Pakistan exceeded 0.67 million tons. However, apple plants are prone to various diseases that can affect the quality as well as the quantity of apples produced. These diseases range from fungal infections to viruses, nematodes, and bacteria, among others.

Diseases such as scab, complex, rust and frog eye leaf spot inhibit the sound development of the apple industry. Among these, scab, which is caused by the ascomycete Venturia inaequalis [3], is the most significant disease. Identifying the area of the leaf affected by scab is quite challenging, and this disease can inflict more damage on the plant than any other disease. Rust disease not only causes serious damage to the apple tree but also leads to reduction of the size of the fruit. Caused by the fungal pathogem, frogeye leaf spot is quite prevalent disease in apple trees [4]. Frog eye leaf spot can lead to fruit infections, severely impacting the apple crop harvest. Therefore, it is essential to accurately diagnose and treat apple diseases to ensure a healthy and productive harvest.

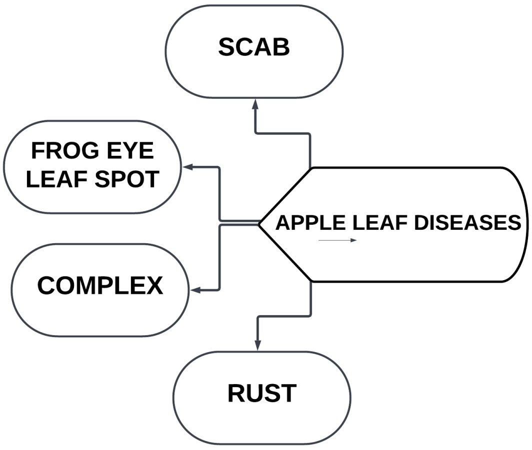


Figure 1.1: Apple Leaf Diseases.

1.1.2 Problem Statement

On time identification and diagnosis of these disease is essential for preventing the economic losses and preserving the nutritional quality of the apples. Timely diagnosis of the correct disease, when it appears, and taking appropriate precautionary measures can help farmers save both the apple harvest and the environment.

For a long time, humans were the primary agents for disease identification. This diagnostic and recognition approach is biased, costly, lengthy, and susceptible to errors. Furthermore, the emergence of new diseases in previously uncharted regions—where there isn't the requisite local expertise for treatment—is another challenge. This backdrop highlights the pressing need for automated methods to detect citrus plant diseases. Thanks to modern tools and rapid computer-aided techniques, real-time scanning and automatic detection of plant anomalies have become more feasible. Traditional machine learning methodologies have demonstrated considerable efficacy in identifying and diagnosing plant ailments, but they primarily cater to a series of image processing tasks such as image segmentation through clustering methods, feature extraction, and utilizing classifiers like Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and

Artificial Neural Network (ANN). However, pinpointing and extracting the most potent pathogenic features is a challenging endeavor, necessitating the expertise of seasoned professionals. This not only introduces variability but also isn't economically efficient in terms of human and financial resources. Contrarily, deep learning has the prowess to innately discern the layered features of diseases, eliminating the need for manual intervention in feature extraction and classification processes.

With the advent of AlexNet [5], the CNNs became the natural choice for the vision applications, leading to the development of high performing models having extensive connections and sophisticated form of convolutions. The recent success of CNNs models was largely due to stacking large number of layers and training very deep networks. CNNs demonstrated significant success in agriculture industry, particularly plant leaf disease detection tasks. [6–9] employed deep ResNet-like architectures with certain modifications for the identification and classification of plant leaf diseases. Due to their substantial model size and intricate network structures, these models achieved remarkable success. Successive works [10–14] focused on developing models that were computationally efficient and could be deployed on resource-constrained devices. Despite being resource efficient, these models fell short of achieving the desired results.

Recently, the transformer architectures have attracted the attention because of their superior performance and ability to model long range dependencies. Equipped with MHSA (multi head self attention), VIT (vision transformers) have achieved valuable results in the field of image classification [15], video classification [16–18], semantic segmentation [19], object detection [20, 21], video object segmentation [22], 3D object detection [23]. However, VITs, due to excessive number of model parameters and high floating-point operations per second (FLOPs), may not be incompatible for real world applications, especially when they need to be deployed on lightweight devices.

Therefore, considering the constraints on storage and computational capacity in embedded device deployment, it is imperative to reduce the excessive number of model parameters and flops of model without compromising on the model's performance.

1.1.3 Objectives

Based on the features of CONV-NET and multi-head self-attention, this paper proposes lightweight apple leaf disease detection and classification model dubbed as AP-

PVIT. Embodied with the characteristics of CNNs and VITs, APPVIT surpasses best lightweight CNNs i.e. EfficientNet in terms of accuracy-efficiency trade-off on lightweight devices. To verify the effectiveness of our model, we compare the results our model with some state of the art CNN architectures. The purpose of this research is to efficiently and accurately identify apple leaf disease in real settings.

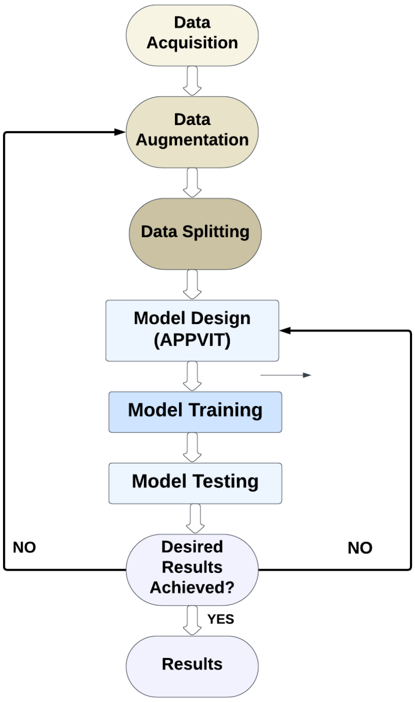


Figure 1.2: Model Architecture.

Research Contributions: The main contributions of this research are as follows:

a) We present a lightweight leaf disease detection network for the identification and classification of five types of apple leaf diseases.

b) Through our findings, we prove that our model outperforms some state of the art CNN architectures in terms of accuracy-efficiency trade-off on lightweight devices.

1.1.4 Thesis Overview

The article unfolds as follows: Chapter 2 delves into the relevant literature. Chapter 3 elucidates the proposed methodology and its execution. Chapter 4 presents the findings from our experiments. Chapter 5 concludes the study, and finally, Chapter 6 charts out potential future directions.

Chapter 2

Literature Review

2.1 Literature Review

Traditionally, the time-consuming task of leaf disease identification was performed by experts in the field: farmers. This conventional approach is time-taking and prone to errors, as the affected areas on leaves are minute and may go unnoticed. Furthermore, people undertaking these jobs were required to possess a thorough understanding of the various diseases, their symptoms, and their respective treatments. An alternative to carry out these tasks manually is to design such systems that are capable of performing these tasks with increased precision and reduced human intervention. Previously, researchers used different machine learning algorithms for the tasks of leaf disease detection. Some of these popular algorithms were support vector machines (SVM) [24–26], random forest [27], filter segmentation [28], K-means clustering [29], KNN [30] and some other image processing methods [31–33]. However, the machine learning algorithms discussed have limited scope and have not achieved the anticipated performance level.

The use of deep learning algorithms in the agriculture industry, especially in plant leaf disease detection, have shown promising results. Since the emergence of AlexNet [5], CNNs have become the de facto choice for the vision researchers, especially in object recognition and classification tasks. CNNs have proven to be an effective contender for plant leaf disease detection, as it bypasses the cumbersome process of segmentation and feature extraction. [34–36] employed CNNs for the classification and identification of different leaf diseases. The ability of CNNs to extract local features from the neighbor pixels enabled them to achieve favorable results when compared with machine learning algorithms. [37] proposed a method combining the features of VGG and InceptionNet for the purpose of apple leaf disease recognition. The model achieved accuracy of 97.14 % on ALDD data set containing 26,377 images of five different diseases.

In [7], the researchers carried out apple leaf disease detection using ResNet-18 and ResNet-34 architectures. Owing to their large size and huge number of parameters, the models achieved accuracies of 99% and 97%. Although these models achieved desirable results, they were not feasible for deployment on lightweight devices. Another successful research carried out in apple leaf disease detection domain is by [38]. The authors trained DenseNet-121 on limited dataset that encompassed only three diseases, segmented into five categories based on disease severity. Their trained model with around 8 million parameters achieved accuracy of 93.1%. In [8], the authors of trained ResNet-197 model for the detection and classification of apple leaf disease. The authors have shown that the model achieves 99.58% accuracy on self collected data set. Although the model is quite successful in disease identification and classification task, its large number of parameters and flops makes it unsuitable for deployment on resource-constrained devices. The authors of [39], combine EfficientNet-B3 and EfficientNet-B4 with certain modifications to built a new model called AFD-Net. The model was trained on plant pathology 2021 - FGVC8 data set and achieved 92.6% test accuracy with 28.2 million parameters.

In [40] the authors substitute all the dense layers from the conventional deep convolutional neural network with a global average pooling layer. An enhanced Softmax classifier is then employed to classify apple leaf disease types. Experimental outcomes on the apple leaf disease image dataset reveal that this technique effectively addresses the overfitting issue, elevates the disease identification accuracy, and substantially cuts down both the training and classification durations of the model. [9] proposed a improved residual network for the identification and classification of apple leaf diseases. To enhance the model's performance, the authors made following changes to ResNet architecture: decomposition of the kernel size, improvements in identity mapping, reduction in residual modules and replacement of the batch normalization (BN) layer. Owing to the changes they made, the model achieved test accuracy of 95.7% with only 3.87 million parameters. Compared to the previous studies, this model achieved significant reduction in model parameters without compromising on model accuracy.

STUDY	CLASSIFICATION METHOD	DATASET	ACCURACY
P. Jiang [37]	VGG and InceptionNet	ALDD Dataset	97.14%
Li and Rai [7]	ResNet-18 and ResNet-34	self-made	99%, 97%
Zhong and Zhao [38]	DenseNet-121	self-made	93.1%
Yu et al. [41]	ROI-aware DCNN	self-made	84.3%
Liu et al. [35]	Improved AlexNet	self-made	97.62%
Pandian et al. [8]	ResNet-197	self-made	99.58%
Yadav et al. [39]	AFD-Net	Plant Pathology 2021	92.6%
Yu et al.. [9]	Modified ResNet	Plant Pathology 2021	95.7%

Table 2.1: Studies in apple leaf diseases identification.

Chapter 3

Dataset and Methodology

3.1 Dataset

In this section, we first discuss the dataset, data augmentation methods, and dataset distribution.

3.1.1 Dataset

The image dataset is taken from Kaggle Plant Pathology 2021 - FGVC8 [42]. The data set comprises 18,632 images belonging to 12 different diseases. These images were taken with diverse backgrounds under natural conditions. The classes are: scab, healthy, frog eye leaf spot, complex and rust. The data distribution is shown in Figure 3.

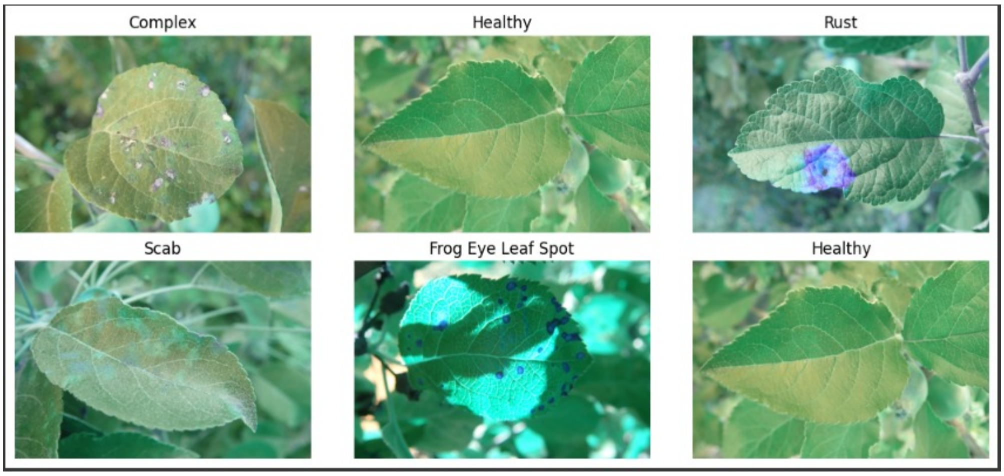


Figure 3.1: Samples of Apple Leaf Diseases.

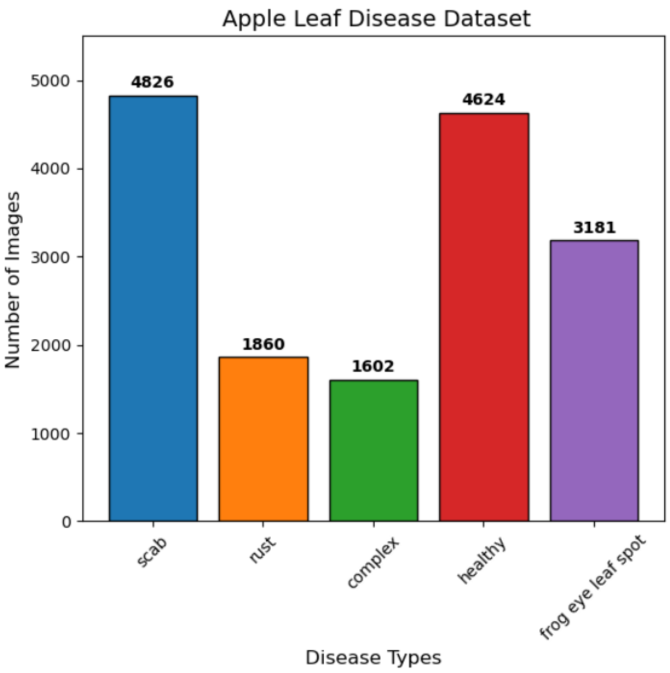


Figure 3.2: Apple Leaf Diseases Dataset before Augmentation.

3.1.2 Data Augmentation

The performance of deep learning problems in closely linked to quality, quantity and relevancy of training data. Insufficient data is one of the biggest challenges when implementing deep learning. Data augmentations are imperative for enhancing and expanding the training dataset as well as in increasing the gener- alizability of deep learning models. It also assists in preventing overfitting.

Data augmentation has been performed using the Keras pre- processing layers API. Keras provides a wide number of data augmentation techniques. The ones we used are:

Random Rotation: Randomly rotates (clockwise or anti- clockwise) the image upto certain degree.

Random Flip: This layer randomly flips the image. Both horizontal and vertical flips are performed by this technique.

Random Translation: This preprocessing layer randomly translates the image.

Brightness: Add brightness to images by certain amount.

Random Zoom: This preprocessing layer randomly zooms (in or out) images.

Random Contrast: The preprocessing layer that randomly adjust contrast.

DISEASES	ORIGINAL IMAGES	AUGMENTATED IMAGES
Healthy	2875	5790
Scab	1116	5580
Rust	962	5772
Complex	2774	5548
Frog Eye Leaf Spot	1908	5724

Table 3.1: Augmented Apple Leaf Disease Dataset.

After applying the data augmentation, the data set comprises of total 28,414 images belonging to five different diseases. A sufficient data set not only help the model to generalize well but also prevents overfitting.

3.1.3 Dataset Distribution

The data set comprising 28,414 images has been split into three sets, namely: training, validation and test set. The split- ting ratio for training, validation and test set is 3:1:1.

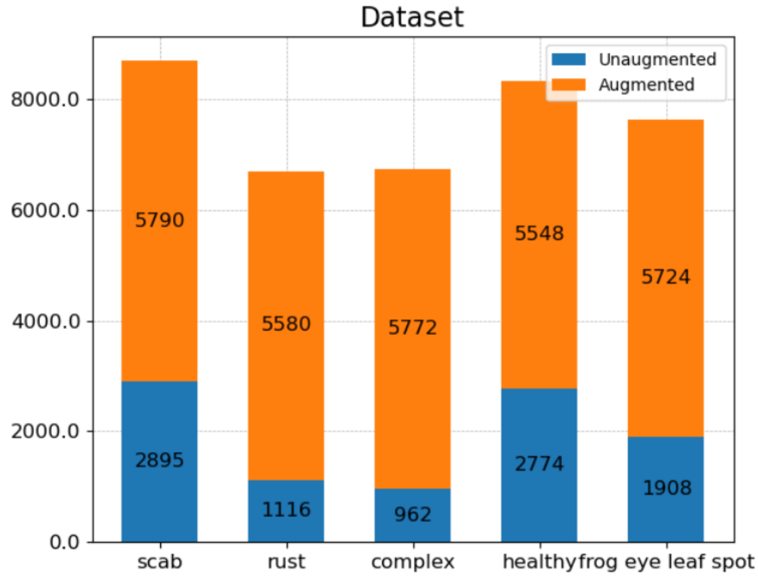


Figure 3.3: Dataset after augmentation.

Figure 3.1 indicates that the dataset is imbalanced, the high- est number of images in dataset are that of Scab class and the lowest number of images are that of Complex class. In or- der to enlarge the dataset to improve the model's performance, data augmentation i.e. random-rotation, random-flip, random- translation, brightness, random-zoom and random-contrast has been performed on dataset to increase the size of the dataset as well as to enhance the quality of the images.

Figure 3.2 and Table 3.1 shows the distribution of dataset after augmentation. After augmentation the dataset now contains 38,069 im- ages compared to 9,655 images before augmentation. The total number of images for each class is as follows follows: Scab 2,895; Rust 6,696; Complex 6,734; Healthy 8,322 and Frog Eye Leaf Spot 7,632.

3.2 Methodology

Before delving into our model, we begin by giving a concise overview of vision transformer and how they can be optimized for lightweight devices. A 2D image fed into VIT is first split into flattened sequence of patches, where H, W represents the height and width of input image and C denote the channel dimensions of the input image [15]. This patchification is performed with a reduction ratio of, where N = HW/P2 is the resulting number of patches and (P, P) is the resolution of each image patch [15]. Before adding position embeddings to the patch embeddings, a special class token (cls) is added to the visual tokens. The resulting sequence of embedding vectors are fed as input to the transformer encoder. The transformer encoder is combination of alternating layers of multi-head self attention (MHSA) and a feed forward networks (FFN).

Multi Head Self Attention: Multi-Head Self Attention (MHSA) mechanism allows the network to model long-term dependencies and learn complex features. Self-attention has been shown to be very effective for learning the global context or long-range spatial dependency of an image, which is critical for visual recognition.

By having multiple heads, the model can capture various types of relationships and dependencies in the data. Multi-head attention typically results in richer representations of the input, making it beneficial for downstream tasks.

Global Average Pooling (GAP): GAP is a layer used within convolutional neural networks, primarily towards the end of the network architecture. Its primary goal is to reduce the spatial dimensions of a 3D tensor (height, width, depth) to a 1D tensor, retaining the depth (number of channels or feature maps). In recent years, GAP has gained popularity in several state-of-the-art architectures, as it provides a lightweight and efficient means to transition from spatial feature extraction to classification decisions.

3.2.1 Overview Architecture of APPVIT

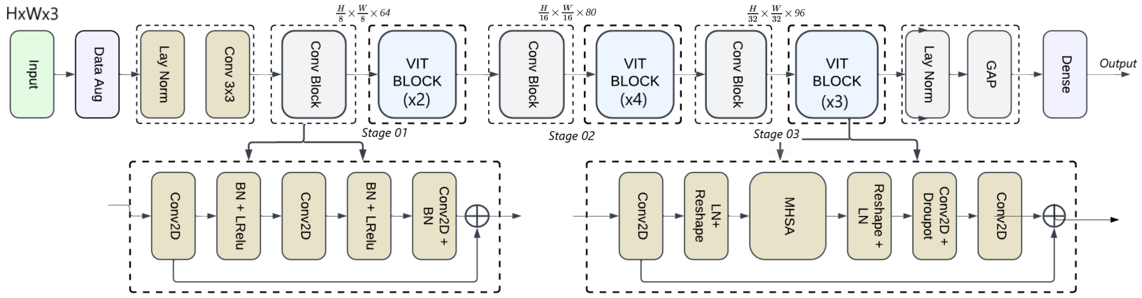


Figure 3.4: Model Architecture.

The aim of this research is to create a hybrid model that is not just lightweight (suitable for deployment on mobile de- vices) but also exhibits strong performance in identification and classification of apple leaf disease dataset. We demonstrate that combining the convolution blocks with multi-head self attention module in hierarchical structure allows the network to model long range dependencies and to learn complex features. Specifically, the network starts with convolution block with VIT blocks stacked afterwards. First, input image is passed through a convolution stem with a kernel size of 3 x 3 and stride 2. APPVIT consists of three sections, each section comprising of convolution block in combination with successive VIT blocks. Convolution block helps the network to encode local features and to narrow down the size of the attention maps. The convolution block is combination of alternating layers of convolution and batch normalization (BN). Afterwards, successive VIT blocks are stacked allowing the network to model global dependencies and contextual information of the features maps. VIT block consist on 2D convolution, multi-head self-attention, layer normalization and droupout layers. From Figure 3.4, it can be seen that 1st section contains 2 successive VIT blocks while 2nd and 3rd section section contains 4 and 3 successive VIT blocks respectively.

The loss function used is Categorical Cross Entropy. Categorical Cross Entropy is also known as Softmax Loss. It's a softmax activation plus a Cross-Entropy loss used for multiclass classification.

Chapter 4

Results and Discussions

4.1 Results

Implementation Details. APPVIT is trained and the experiments are conducted on plant pathology 2021 - FGVC8 dataset. The models is built with Tensorflow 2.11, and trained from scratch for 60 epochs on NVIDIA TESLA P100 GPUs using AdamW [43] optimizer and cosine learning rate scheduler. The batch size was set to 32. The input images are resized and randomly cropped into 256 × 256. The initial learning rate is 0.002 with weight decay of 0.000125. The data augmentation performed on dataset is discussed in detail in the above dataset section.

Parameter	Values
Learning rate	0.002
Batch-size	32
Optimizer	AdamW
Epoch	65

Table 4.1: Setting of model parameters.

Results.

This section discusses the performance of the proposed APPVIT on plant pathology 2021 - FGVC8 dataset. These results are obtained using standard performance metrics. Precision, recall and F-score are some of the common performance metrics used to measure the performance of the classification model. The formulae for the performance metrics i.e. Accuracy, precision, recall and F-measure are given below:

$$Accuaracy = \frac{No.\, of\, images\, correctly\, classified}{Total\, no.\, of\, images} \tag{4.1.1}$$

$$Precision = \frac{All\, TP}{TP\, +\, FP} \tag{4.1.2}$$

$$Recall = \frac{All\, TP}{TP\, +\, FN} \tag{4.1.3}$$

$$F - measure = \frac{2\, *\, Precision\, *\, Recall}{Precision\, +\, Recall} \tag{4.1.4}$$

TP stands for True Positive, FN stands for false negative and FP stands for false positive.

A true positive occurs when the model accurately identifies a case as belonging to the positive class. Conversely, a true negative arises when the model correctly classifies a case as not belonging to the positive class.

A false positive happens when the model wrongly categorizes a case as part of the positive class. On the other hand, a false negative occurs when the model mistakenly classifies a positive case as being negative.

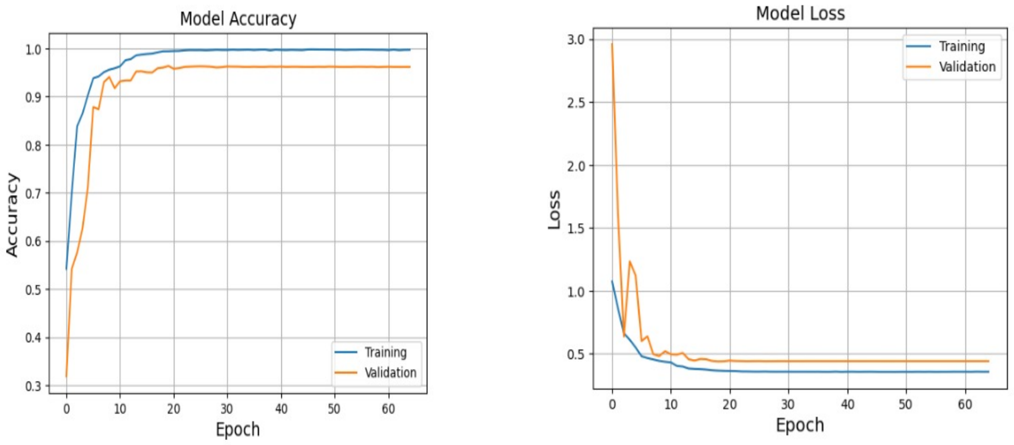
In addition, this research also takes into consideration the relation of parameter memory, number of parameters and FLOPs with the accuracy of the model.

Figure 4.1 shows the training and validation accuracy and loss curves for our proposed model: APPVIT.

Table 4.2 presents the comparison of the APPVIT with some state of the art architectures. The comparisons here involve three factors: the accuracy, number of Parameter and flops. APPVIT achieves accuracy of 96.4% on plant pathology 2021 - FGVC8 dataset. Compared to the state of the art CNN based architectures, APPVIT surpass all the models in terms of not just accuracy but also in terms on Parameter Memory / Mb as well.

Comparisons with ResNets, Inception and VGG.

We first compare APPVIT with ResNet models, such as ResNet-152 and ResNet-50. Specifically. compared to ResNet-152, APPVIT achieves 16.4% better top-1 accuracy,



(a) Training and Validation Accuracy.

(b) Training and Validation Loss.

Figure 4.1: Training and Validation Curves for Accuracy and Loss.

and is remarkably more efficient, using 97.8% fewer parameters and 94.35% flops. Compared to state of the art ResNet-50, APPVIT obtains 11.2% higher accuracy and yet has 94.92% fewer parameters and 83.9% fewer flops. When compared to Inception-V3, APPVIT obtains 16.1% better accuracy with 88.73% fewer flops and 94.54% fewer parameters. Moreover, APPVIT achieves 13.4% higher accuracy when compared VGGNet. Our model has staggering 99.06% fewer parameters and 95.79% fewer flops than VGGNet.

In terms of parameter memory / mb, our model APPVIT exhibited reduction of 99.05%, 97.81% and 94.89% when compared to VGGNet, ResNet-152 and ResNet-50, respectively.

Comparison with Efficient CNNs.

We also compare our model with lightweight and efficient CNNs. When compared to EfficientNet-B3 and EfficientNet-B4 our model achieves 4.3% and 5% better top-1 accuracy. In terms of number of parameters, flops and parameter memory / mb, our model demonstrated reduction of 64.22%, 89.17% and 89.12% respectively, in comparison to EfficientNet-B3. In comparison to EfficientNet-B4, our model has 84.67% fewer flops, 93.16% fewer parameters and 93.10% reduced parameter memory / mb.

Comparisons with AFD-Net and MSO-ResNet.

When compared to AFD-Net [39], our proposed model achieves 3.8% better accuracy. Moreover, the APPVIT has 95.39% fewer parameters than AFD-Net and is more com-

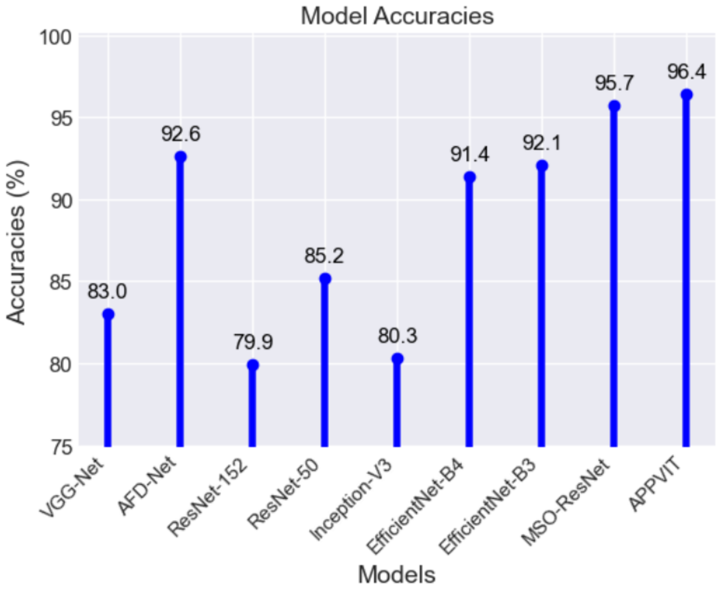


Figure 4.2: Comparison of Model Accuracies.

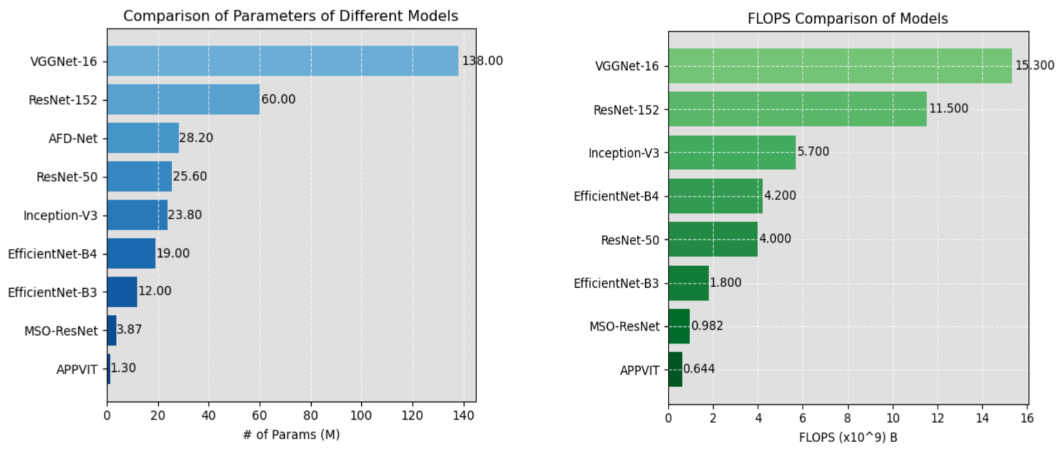
Method	#Param (M)	FLOPS (B)	Parameter Memory / Mb	Accuracy
VGGNet	138	15.3	525	83%
ResNet-152	60	11.4	228	79.9%
AFD-Net	28.2	-	-	92.6%
ResNet-50	25.6	4	97.66	85.2%
Inception-V3	23.8	5.7	90.5	80.3%
EfficientNet-B3	12	1.8	45.8	92.1%
EfficientNet-B4	19	4.2	72.33	91.4%
MSO-ResNet	3.87	0.982	14.77	95.7%
APPVIT	1.3	0.644	4.99	96.4%

Table 4.2: Performance comparison of different experiment schemes.

patible for deployment on lightweight devices.

In comparison to MSO-ResNet [9], our model saw 66.41% reduction in number of parameters and 34.38% fewer flops. Our model despite being lightweight achieves 0.7% better accuracy than MSO-ResNet.

In evaluating various deep learning models for a particular task, the performance metrics of precision, recall, and F1 score serve as critical indicators of each model's capability.



(a) Comparison of different Model Parameters . (b) Comparison of different Model FLOPS.

Figure 4.3: Comparison of APPVIT with State of the Art Architectures.

Among the models analyzed, APPVIT showcased superior results, indicating its robustness and accuracy in the given context.

Specifically, APPVIT achieved a precision of 0.967, a recall of 0.959, and an F1 score of 0.963. These numbers outpace those of other popular architectures. For instance, ResNet variants, such as ResNet-34, ResNet-50, and ResNet-152, recorded F1 scores of 0.888, 0.897, and 0.886 respectively, all of which are noticeably lower than APPVIT's score.

DenseNet121 and EfficientNetb0, though closer contenders with F1 scores of 0.916 and 0.920 respectively, still lag behind the performance exhibited by APPVIT. Even the MSO-ResNet, which garnered a commendable F1 score of 0.957, doesn't quite match the prowess of APPVIT.

Based on the provided performance metrics in Table 4.3, APPVIT emerges as the clear front-runner. Its higher precision indicates a better ability to correctly identify positive samples, and its recall suggests fewer false negatives. The combined effect of these is reflected in its leading F1 score, demonstrating a balanced and superior performance in comparison to the other models.

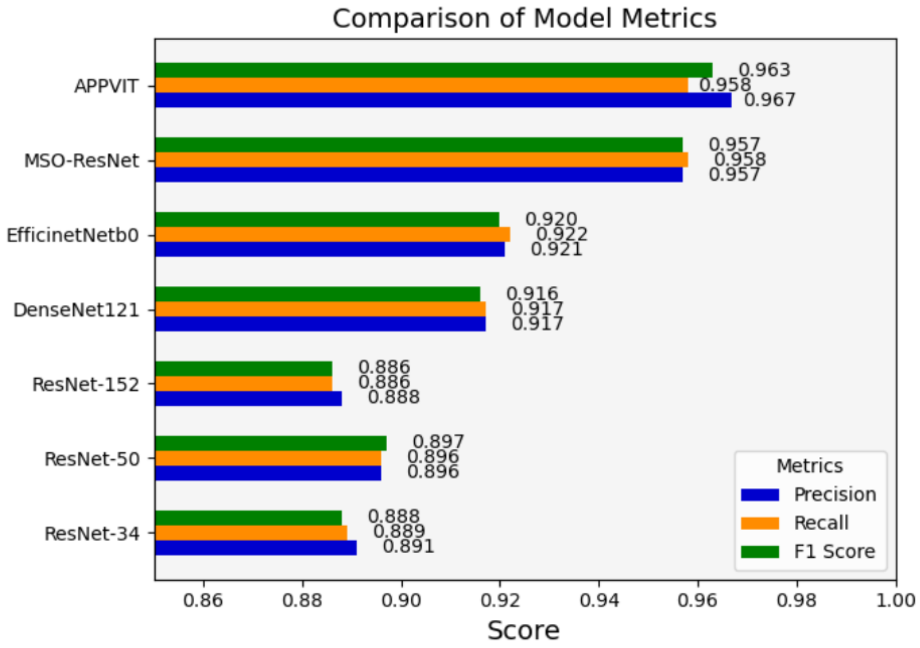


Figure 4.4: Comparison of Model Accuracies.

Model	Precision	Recall	F1 Score
ResNet-34	0.891	0.889	0.888
ResNet-50	0.896	0.896	89.70
ResNet-152	0.888	0.886	0.886
DenseNet121	0.917	0.917	0.916
EfficinetNetb0	0.921	0.922	0.920
MSO-ResNet	0.957	0.958	0.957
APPVIT	0.967	0.959	0.963

Table 4.3: Performance comparison of different Models.

Chapter 5

Conclusion

In this work, we propose a lightweight apple leaf disease detection model dubbed as APPVIT from the perspective of on-device deployment. Embodied with the characteristics of CNNs and VITs, APPVIT achieves superior results on visual recognition task, when compared to some of the state of the art CNN architectures. Extensive experiments on plant pathology 2021 - FGVC8 shows that APPVIT outperforms ResNets and their modified versions while being faster than them. We have successfully proven that not only our model achieves highest accuracy compared to the other state of the art architectures but also it has much fewer number of parameters, flops and parameter memory / mb compared to these state of the art architectures. These experiments verify both the effectiveness and efficiency of our model. The detection system we have developed can be very helpful tool for farmers and apple growers to aid them in diagnosis, quantification and follow-up of infections.

Chapter 6

Future Work

This work is based on five types of apple leaf diseases. To conduct a more in-depth analysis, other apple leaf diseases classes can be explored. Absence of data has remained the biggest hurdle while conducting this research. In future, the data of other types of apple disease can be added.

Furthermore, in future, such lightweight models can be extended to other fruits and vegetables as well.

Next challenge that we are dealing with is the detection of these diseases during early stages.

References

[1] World Cancer Research Fund International. *Diet, nutrition, physical activity and cancer: a global perspective: a summary of the Third Expert Report*. World Cancer Research Fund International, 2018.

[2] Susanna C Larsson, Jarmo Virtamo, and Alicja Wolk. Total and specific fruit and vegetable consumption and risk of stroke: a prospective study. *Atherosclerosis*, 227 (1):147–152, 2013.

[3] William E MacHardy et al. *Apple scab: biology, epidemiology, and management.* American Phytopathological Society (APS Press), 1996.

[4] Janna L Beckerman. Frogeye leaf spot, Fact Sheet N°282, 2021. `https://fff.hort.purdue.edu/article/frogeye-leaf-spot/`.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[6] Shanwen Zhang, QingQing Zhang, Ping Li, et al. Apple disease identification based on improved deep convolutional neural network. *Journal of Forestry Engineering*, 4(4):107–112, 2019.

[7] Xin Li and Laxmisha Rai. Apple leaf disease identification and classification using resnet models. In *2020 IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT)*, pages 738–742. IEEE, 2020.

[8] Arun Pandian, K Kanchanadevi, NR Rajalakshmi, and G Arulkumaran. An improved deep residual convolutional neural network for plant leaf disease detection. *Computational Intelligence and Neuroscience*, 2022, 2022.

[9] Helong Yu, Xianhe Cheng, Chengcheng Chen, Ali Asghar Heidari, Jiawen Liu, Zhennao Cai, and Huiling Chen. Apple leaf disease recognition method with improved residual network. *Multimedia Tools and Applications*, 81(6):7759–7782, 2022.

[10] Asif Iqbal Khan, SMK Quadri, Saba Banday, and Junaid Latief Shah. Deep diagnosis: A real-time apple leaf disease detection system based on deep learning. *computers and Electronics in Agriculture*, 198:107093, 2022.

[11] Bui Thi Hanh, Hoang Van Manh, and Ngoc-Viet Nguyen. Enhancing the performance of transferred efficientnet models in leaf image-based plant disease classification. *Journal of Plant Diseases and Protection*, 129(3):623–634, 2022.

[12] Fei Gao, Jiming Sa, Zhuoer Wang, and Zhongyu Zhao. Cassava disease detection method based on efficientnet. In *2021 7th International Conference on Systems and Informatics (ICSAI)*, pages 1–6. IEEE, 2021.

[13] Ümit Atila, Murat Uçar, Kemal Akyol, and Emine Uçar. Plant leaf disease classification using efficientnet deep learning model. *Ecological Informatics*, 61:101182, 2021.

[14] Lili Li, Shujuan Zhang, and Bin Wang. Apple leaf disease identification with a small and imbalanced dataset based on lightweight convolutional networks. *Sensors*, 22 (1):173, 2021.

[15] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[16] Haoqi Fan, Bo Xiong, Karttikeya Mangalam, Yanghao Li, Zhicheng Yan, Jitendra Malik, and Christoph Feichtenhofer. Multiscale vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6824–6835, 2021.

[17] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6836–6846, 2021.

[18] Gedas Bertasius, Heng Wang, and Lorenzo Torresani. Is space-time attention all you need for video understanding? In *ICML*, volume 2, page 4, 2021.

[19] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12179–12188, 2021.

[20] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.

[21] Yanghao Li, Hanzi Mao, Ross Girshick, and Kaiming He. Exploring plain vision transformer backbones for object detection. In *European Conference on Computer Vision*, pages 280–296. Springer, 2022.

[22] Brendan Duke, Abdalla Ahmed, Christian Wolf, Parham Aarabi, and Graham W Taylor. Sstvos: Sparse spatiotemporal transformers for video object segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5912–5921, 2021.

[23] Ishan Misra, Rohit Girdhar, and Armand Joulin. An end-to-end transformer model for 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2906–2917, 2021.

[24] Caio Bruno Wetterich, Ruan Felipe de Oliveira Neves, José Belasque, and Luis Gustavo Marcassa. Detection of citrus canker and huanglongbing using fluorescence imaging spectroscopy and support vector machine technique. *Applied optics*, 55(2): 400–407, 2016.

[25] Feng Jiang, Yang Lu, Yu Chen, Di Cai, and Gongfa Li. Image recognition of four rice leaf diseases based on deep learning and support vector machine. *Computers and Electronics in Agriculture*, 179:105824, 2020.

[26] Prabira Kumar Sethy, Nalini Kanta Barpanda, Amiya Kumar Rath, and Santi Kumari Behera. Deep feature based rice leaf disease identification using support vector machine. *Computers and Electronics in Agriculture*, 175:105527, 2020.

[27] Andrzej Wójtowicz, Jan Piekarczyk, Bartosz Czernecki, and Henryk Ratajkiewicz. A random forest model for the classification of wheat and rye leaf rust symptoms

based on pure spectra at leaf scale. *Journal of Photochemistry and Photobiology B: Biology*, 223:112278, 2021.

[28] Radhika Kamath, Mamatha Balachandra, and Srikanth Prabhu. Crop and weed discrimination using laws' texture masks. *International Journal of Agricultural and Biological Engineering*, 13(1):191–197, 2020.

[29] Kai Tian, Jiuhao Li, Jiefeng Zeng, Asenso Evans, and Lina Zhang. Segmentation of tomato leaf images based on adaptive clustering number of k-means algorithm. *Computers and Electronics in Agriculture*, 165:104962, 2019.

[30] Archana Chaudhary, Ramesh Thakur, Savita Kolhe, and Raj Kamal. A particle swarm optimization based ensemble for vegetable crop disease recognition. *Computers and Electronics in Agriculture*, 178:105747, 2020.

[31] Saeed Nosratabadi, Sina Ardabili, Zoltan Lakner, Csaba Mako, and Amir Mosavi. Prediction of food production using machine learning algorithms of multilayer perceptron and anfis. *Agriculture*, 11(5):408, 2021.

[32] Jafar Tavoosi, Chunwei Zhang, Ardashir Mohammadzadeh, Saleh Mobayen, and Amir H Mosavi. Medical image interpolation using recurrent type-2 fuzzy neural network. *Frontiers in Neuroinformatics*, 15:667375, 2021.

[33] Muhammad Zubair Asghar, Ikram Ullah, Shahab Shamshirband, Fazal Masud Kundi, and Ammara Habib. Fuzzy-based sentiment analysis system for analyzing student feedback and satisfaction. 2019.

[34] Alvaro Fuentes, Sook Yoon, Sang Cheol Kim, and Dong Sun Park. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors*, 17(9):2022, 2017.

[35] Bin Liu, Yun Zhang, DongJian He, and Yuxiang Li. Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry*, 10(1):11, 2017.

[36] Keke Zhang, Qiufeng Wu, Anwang Liu, Xiangyan Meng, et al. Can deep learning identify tomato leaf disease? *Advances in multimedia*, 2018, 2018.

[37] Peng Jiang, Yuehan Chen, Bin Liu, Dongjian He, and Chunquan Liang. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access*, 7:59069–59080, 2019.

[38] Yong Zhong and Ming Zhao. Research on deep learning in apple leaf disease recognition. *Computers and Electronics in Agriculture*, 168:105146, 2020.

[39] Anju Yadav, Udit Thakur, Rahul Saxena, Vipin Pal, Vikrant Bhateja, and Jerry Chun-Wei Lin. Afd-net: Apple foliar disease multi classification using deep learning on plant pathology dataset. *Plant and Soil*, 477(1-2):595–611, 2022.

[40] Shanwen Zhang, Subing Zhang, Chuanlei Zhang, Xianfeng Wang, and Yun Shi. Cucumber leaf disease identification with global pooling dilated convolutional neural network. *Computers and Electronics in Agriculture*, 162:422–430, 2019.

[41] Hee-Jin Yu and Chang-Hwan Son. Apple leaf disease identification through region-of-interest-aware deep convolutional neural network. *arXiv preprint arXiv:1903.10356*, 2019.

[42] Kai; Snavely Noah; Belongie Serge; Khan Awais Thapa, Ranjita; Zhang. Plant pathology 2021 - fgvc8, 2021. URL `https://kaggle.com/competitions/plant-pathology-2021-fgvc8`.

[43] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.