

Unveiling Agricultural Insights: Leveraging Deep Learning for Enhanced Diagnostic Accuracy in Maize Disease Detection with Explainable Artificial Intelligence

by

Basit Hussain
21141064

Malika Muradi
21241057

Christian Boateng
22101816

Eliya Christopher Nandi
21341039

Imenagitero Ulysse Tresor
21101333

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
October 2024

© 2024. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Basit Hussain
21141064

Malika Muradi
21241057

Christian Boateng
22101816

Eliya Christopher Nandi
21341039

Imenagitero Ulysse Tresor
21101333

Approval

The thesis titled “Unveiling Agricultural Insights: Leveraging Deep Learning for Enhanced Diagnostic Accuracy in Maize Disease Detection with Explainable Artificial Intelligence” submitted by

1. Basit Hussain (21141064)
2. Malika Muradi (21241057)
3. Christian Boateng (22101816)
4. Eliya Christopher Nandi (21341039)
5. Imenagitero Ulysse Tresor (21101333)

Of Summer, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on Oct 17, 2024.

Examining Committee:

Supervisor:

Annajiat Alim Rasel
Senior Lecturer
Department of Computer Science and Engineering
Brac University

Secondary Supervisor:

Amitabha Chakrabarty, PhD
Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Md. Aquib Azmain
Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

Maize is a vital crop that feeds over a billion people worldwide and supports numerous industries. However, maize production is threatened by devastating plant diseases such as Maize Lethal Necrosis (MLN) and Maize Streak Virus (MSV), which can lead to significant yield losses and economic impacts, particularly in Sub-Saharan Africa. Therefore, to prevent significant losses of this essential crop, farmers need to be equipped with advanced tool that enables accurate and timely disease detection. In this regards, we have implemented a comparative performance analysis of five Transfer Learning (TF) (EfficientNetV2B2, ResNet50, InceptionV3, VGG16, and Xception) and five Vision Transformer (ViT) (SWIN, DaViT, MobileViT, MaxViT, and Involutional Neural Network (INN)) models for maize crop disease detection. We subsequently developed a fusion model that integrates MobileViT and DaViT. Afterward, the performance of the models was evaluated using multiple metrics such as precision, recall, and f1-score. The proposed fusion model perform best across all the metrics with an accuracy of 96.67%, recall of 95.84%, precision of 96.34%, and a f1-score of 96.54%. For transparent decision-making, three explainable artificial intelligence (XAI) techniques such as saliency map, gradient weighted class activation mapping (Grad-CAM), and local interpretable model agnostic explanations (LIME) have been implemented. Finally, we deployed the proposed fusion model on a Raspberry Pi to facilitate real-time detection of maize diseases.

Keywords: maize diseases detection, transfer learning, vision transformers, fusion model, hardware deployment, Grad-CAM, XAI, LIME, saliency map;

Table of Contents

Declaration	i
Approval	ii
Abstract	iv
Table of Contents	v
List of Figures	vii
List of Tables	ix
Nomenclature	xi
1 Introduction	1
1.1 Background and Motivation	1
1.2 Problem Statement	3
1.3 Aims and Objectives	4
1.4 Contributions	4
1.5 Structure of The Paper	5
2 Related Work	6
3 Methodology	21
3.1 Work Plan	21
3.1.1 Data Collection	22
3.1.2 Data Preprocessing	23
3.2 Transfer Learning Models	24
3.2.1 ResNet50	24
3.2.2 InceptionV3	25
3.2.3 VGG16	26
3.2.4 EfficientNetV2B2	27
3.2.5 Xception	27
3.3 Vision Transformer Models and Hybrid Model	28
3.3.1 Shifted Window Transformer (SWIN)	29
3.3.2 Dual-Attention Vision Transformer (DaViT)	30
3.3.3 MobileViT	31
3.3.4 MaxViT	32
3.3.5 Involutional Neural Network	33
3.4 Proposed Fusion Model	34

4	Results and Discussion	35
4.1	Transfer Learning Model’s Performance	35
4.1.1	ResNet50	36
4.1.2	InceptionV3	37
4.1.3	VGG16	37
4.1.4	EfficientNetV2B2	38
4.1.5	Xception	39
4.2	Vision Transformer & Hybrid Model’s Performance	40
4.2.1	SWIN ViT	41
4.2.2	DaViT	42
4.2.3	MobileViT	44
4.2.4	MaxViT	46
4.2.5	INN	48
4.2.6	Fusion Model	50
4.3	Comparative Analysis	52
5	Hardware Deployment	56
5.1	Setup	57
5.2	Model Compression and Deployment	58
5.3	System Architecture	58
5.4	System Demonstration	59
5.5	Inference Results	59
6	Explainable Artificial Intelligence (XAI)	63
6.1	Saliency Map	63
6.2	Grad-CAM	65
6.3	LIME	67
7	Conclusion and Future Direction	70
7.1	Conclusion	70
7.2	Future Direction	70
	Bibliography	76

List of Figures

1.1	Samples of maize leaves infected vs healthy	1
1.2	Organization of the Paper.	5
3.1	Diagram of the Work Plan.	22
3.2	Sample images.	23
3.3	Distribution of Images in Each Class.	23
3.4	ResNet50 Model Architecture.	25
3.5	Skip Connection.	25
3.6	VGG16 Architecture.	26
3.7	EfficientNetV2 Structures.	27
3.8	ViT Structures.	28
3.9	Swin Transformer vs ViT	30
3.10	DaViT Architecture	31
3.11	MobileViT Architecture	32
3.12	MaxViT Architecture	33
3.13	INNs Architecture	34
3.14	Fusion Model Structure	34
4.1	ResNet50 Model Confusion Matrix	36
4.2	InceptionV3 Model Accuracy & Loss Metrics	37
4.3	VGG16 Model Accuracy & Loss Metrics	38
4.4	EfficientNetV2B2 Model Confusion Matrix	39
4.5	Xception Model Accuracy Metric	39
4.6	Xception Model Loss Metric	40
4.7	Swin ViT Model Confusion Matrix	41
4.8	SwinViT Accuracy Metric	42
4.9	SwinViT Loss Metric	42
4.10	DaViT Model Confusion Matrix	43
4.11	DaViT Model Metrics	44
4.12	MobileViT Model Confusion Matrix	44
4.13	MobileViT Accuracy Metric	45
4.14	MobileViT Loss Metric	46
4.15	MaxViT Model Confusion Matrix	46
4.16	MaxViT Accuracy Metric	47
4.17	MaxViT Loss Metric	48
4.18	INN Model Confusion Matrix	49
4.19	INN Model Metric	49
4.20	Fusion Model Confusion Matrix	50
4.21	Accuracy Metrics for Fusion Model	51

4.22	Loss Metrics for Fusion Model	51
4.23	All Models Comparison	52
4.24	Training time Vs Memory usage	55
5.1	Raspberry Pi Setup	57
5.2	Model Deployment	58
5.3	Hardware System Architecture	59
5.4	System Demonstration	61
5.5	Inference performance Metrics	62
6.1	Saliency map results for a healthy maize crop using fusion model: The original image (left), saliency map (center), and saliency map with heatmap overlay (right).	64
6.2	Saliency map results for a MLN maize crop using fusion model: The original image (left), saliency map (center), and saliency map with heatmap overlay (right).	64
6.3	Saliency map results for a MSV_1 maize crop using fusion model: The original image (left), saliency map (center), and saliency map with heatmap overlay (right).	65
6.4	Saliency map results for an MSV_2 maize crop using the fusion model: the original image (left), the saliency map (center), and the saliency map with heatmap overlay (right).	65
6.5	Grad-CAM results for a healthy maize crop using fusion model: the original image (left), Grad-CAM (center), and Grad-CAM with heatmap (right).	66
6.6	Grad-CAM results for a MLN maize crop using fusion model: the orig- inal image (left), GradCAM (center), and Grad-CAM with heatmap (right).	66
6.7	Grad-CAM results for a MSV_1 maize crop using fusion model: the original image (left), Grad-CAM (center), and Grad-CAM with heatmap (right).	67
6.8	Grad-CAM results for a MSV_2 maize crop using fusion model: the original image (left), Grad-CAM (center), and Grad-CAM with heatmap (right).	67
6.9	LIME results for a health maize crop using fusion model: the original image (left), LIME mask (center), and LIME with heatmap (right).	68
6.10	LIME results for a mln maize crop using fusion model: the original image (left), LIME mask (center), and LIME with heatmap (right).	68
6.11	LIME results for a msv_1 maize crop using fusion model: the original image (left), LIME mask (center), and LIME with heatmap (right)	69
6.12	LIME results for a msv_2 maize crop using fusion model: the original image (left), LIME mask (center), and LIME with heatmap (right).	69

List of Tables

2.1	Summary Table For Selected Papers	14
2.3	Comparison of Different Papers With Our Paper	20
4.1	Parameter Settings for Different Models	35
4.2	ResNet50 Model Report.	36
4.3	InceptionV3 Model Report.	37
4.4	VGG16 Model Report.	38
4.5	EfficientNetV2B2 Model Report.	38
4.6	Xception Model Report.	40
4.7	ViT HyperParameter Settings	41
4.8	Swin Transformer Model Report.	42
4.9	DaViT Model Report.	43
4.10	MobileViT Model Report.	45
4.11	MaxViT Model Report.	47
4.12	INN Model Report	48
4.13	Fusion Model Report	50
4.14	Transfer Learning vs. Vision Transformers	53
4.15	Selection Criterias for Vision Transformer(ViT) Models	53
5.1	Raspberry Pi Specifications	56
5.2	Hardware and Software Requirements	57
5.3	Inference Performance Metrics for Each Class	60

Nomenclature

This section provides a list of abbreviations and their full forms to assist you in navigating the document and comprehending the technical terms and acronyms used.

BiLSTM Bidirectional Long Short-Term Memory

CARAFE Content-Aware ReAssembly of FEatures

CENet Cascaded Encoder Network

CNN Convolutional Neural Network

DaViT Dual Attention Vision Transformer

EfficientNetB0 EfficientNet with baseline model B0

EfficientNetV2B2 Efficient Convolutional Neural Network Model (Version V2B2)

Grad – CAM Gradient-weighted Class Activation Mapping

InceptionV3 Inception Version 3 Convolutional Neural Network

LEMoxinet Ensemble of Xception and MobileNet

LIME Local Interpretable Model-agnostic Explanations

LS – RCNN Large Scale Region-based Convolutional Neural Network

MaxViT Max Pooling Vision Transformer

MCMV Maize Chlorotic Mottle Virus

MFasterR – CNN Modified Faster Region-based Convolutional Neural Network

MLN Maize Lethal Necrosis

Mo – BioNetV2 MobileNet Version 2

MobileNet Mobile-first Convolutional Neural Networks

MSV Maize Streak Virus

PReLU Parametric Rectified Linear Unit

ResNet50 Residual Network with 50 layers

SCMV Sugarcane Mosaic Virus

SVM Support Vector Machine

Swin Shifted Window Transformer

TF Transfer Learning

VGG16 Visual Geometry Group with 16 layers

ViT Vision Transformer

WT – GBF Wavelet Threshold-guided Bilateral Filtering

XAI Explainable Artificial Intelligence

Xception Extreme Inception Model (CNN Architecture)

Chapter 1

Introduction

1.1 Background and Motivation

Agriculture plays a critical role in maintaining global food security and economic stability [1]. Among the staple crops, maize scientifically known as *Zea mays* and commonly referred to as corn, forms the backbone of human diets. In Sub-Saharan Africa, maize is widely cultivated, particularly in Tanzania, where it spans over 5 million hectares, with an average annual consumption of 128 kg per person. This versatile crop is consumed directly as corn on the cob and is also processed into various products like cornmeal, flour, tortillas, snacks, sweeteners, and cornstarch. Additionally, maize is essential for livestock feed, biofuel production, and biodegradable plastics, making it a remarkably versatile crop [19].



Figure 1.1: Samples of maize leaves infected vs healthy

The global significance of maize is evident in its extensive cultivation and diverse application, with annual consumption exceeding 1.2 billion metric tons. As the world's most produced grain, it accounts for approximately 30% of total cereal production, meeting the dietary needs of over 1 billion people worldwide. In low-income and food-deficit countries, maize can contribute up to 60% of daily caloric intake, highlighting its vital role in combating hunger and malnutrition. Its importance extends beyond human consumption, as more than 50% of total maize production is used as animal feed, sustaining the livestock industry. Additionally, the growing biofuel sector consumes about 150 million metric tons of maize annually for ethanol production, illustrating its influence on global energy markets. The versatility and

multi-sectoral demand for maize mean that disruptions in its supply due to diseases could have severe consequences. Economic losses from reduced yields, estimated at \$2 billion annually for maize disease outbreaks in Africa alone, can destabilize local markets and threaten food security. The interconnectedness of maize across food, feed, and energy sectors makes protecting it a priority, as the ripple effects of disease can escalate into higher food prices, livestock feed shortages, and decreased biofuel production, affecting economies on a global scale.

Despite its global importance, maize production faces significant challenges such as Northern Leaf Blight, Common Rust, Gray Leaf Spot, Maize Lethal Necrosis (MLN), and Maize Streak Virus (MSV). Among these, MLN and MSV are particularly devastating, posing significant threats to maize crops, especially in Sub-Saharan Africa. MLN is caused by a combination of Maize Chlorotic Mottle Virus (MCMV) and Sugarcane Mosaic Virus (SCMV), leading to symptoms such as yellowing, leaf necrosis, stunted growth, and eventual plant death. These diseases can cause substantial yield losses, affecting food security and livelihoods in the affected regions. MSV, transmitted by leafhopper insects, manifests through streaking or striping of the leaves, stunted growth, and reduced grain quality. Outbreaks of MLN and MSV have been documented across several African regions, including Kenya's Central and Rift Valley Provinces, Ethiopia, Nigeria, Tanzania, Uganda, and more recently, Rwanda [53].

Based on the recent U.S government agriculture statistics [56], Africa's maize production reached nearly 91 million metric tons in the 2023/2024 trade year. However, forecasts indicate a potential decline to 88 million metric tons in 2024/2025. The anticipated drop is driven by several challenges, including disease outbreaks, which could severely affect the continent's agricultural productivity and food security. Accurate and timely disease diagnosis is crucial for effective treatment that ensures food security, and maximizes the crop yields. Unfortunately, limited access to advanced diagnostic technologies in rural farming communities exacerbates the problem, delaying interventions and resulting in widespread crop damage. Traditional diagnostic methods that rely on expert field assessments are often time-consuming, sometimes inaccurate, and inaccessible to smallholder farmers, highlighting the need for more efficient and accurate approaches.

Recent research has leveraged advancements in artificial intelligence, particularly deep learning techniques, to classify maize diseases. Convolutional Neural Networks (CNNs) have demonstrated superior performance in disease diagnosis compared to traditional machine learning models. CNN-based models such as ResNet50, VGG16, MobileNet, and MaizeNet have shown promise in this area. Despite this progress, there remains a lack of reliable studies specifically addressing the diagnosis of MLN and MSV—two diseases that are often diagnosed late, resulting in critical consequences for farmers and stakeholders.

To bridge this gap, we implemented various transformer learning models and vision transformers and achieved high-accuracy in maize disease detection using state-of-the-art AI models. While models such as MobileViT have performed well from Vision transfer models, we took an additional step and built a fusion model, with

impressive performance. This fusion Transformer model is lightweight and suitable for implementation on any device.

Moreover, there is a noticeable gap in the application of Explainable AI (XAI) techniques in these studies. While these offer transparency and insights into the decision-making processes of diagnostic models, which is essential for building trust and understanding among non-expert users, specifically farmers. In this regard, we incorporated XAI to enhance the reliability of the diagnostic tool in practical applications.

The dataset used, sourced from the Nelson Mandela African Institution of Science and Technology and the Tanzania Agricultural Research Institute. It contains a total of 17,277 images across three different classes: healthy, MLN, and MSV. We analyzed these images to identify patterns associated with different maize diseases by implementation of CNN, Vision transformer, and fusion models.

The CNN models used in this study include EfficientNetV2B2, ResNet50, InceptionV3, VGG16, and Xception owing to their proven effectiveness. Additionally, we employed the vision transformers models such as MobileViT, Swin, Davit, MaxViT, and Convolutional Neural Network to explore their potential in enhancing accuracy. Lastly, we implemented a fusion model combined the strength of two well performed models to create a highly effective diagnostic tool. The dataset was split into training and validation sets, and all models were trained and evaluated to assess their performance.

By incorporating XAI techniques, we also ensure transparency and provide insights into the decision-making processes of the diagnostic models. This helps farmers and other stakeholders understand the reasoning behind disease identification, fostering trust and promoting the use of these advanced tools.

1.2 Problem Statement

Maize, one of the world's most vital crops, faces a constant threat from diseases, which can drastically reduce crop yields and quality. Diseases have a particularly strong influence on crop yield in places such as Tanzania, where maize agriculture is critical to food security and economic stability. Traditional disease diagnostic methods frequently rely on visual inspection by agricultural professionals, which is time-consuming, subjective, error-prone and sometimes unavailable to smallholder farmers. This highlight the need for a highly accurate and effective diagnostic tool in this field. In this regard, Our study improved diagnostic accuracy using Fusion model and implemented three XAI methodologies, making predictions more readily identifiable and interpretable by all agricultural stakeholders. This research aspires to give farmers a reliable, accessible, and intelligible diagnostic tool, thereby enhancing maize disease management and agricultural output.

1.3 Aims and Objectives

The main objective of our paper is to develop a reliable and transparent method for maize diseases detection, incorporating XAI techniques for better interpretation and detection. Our detailed aims and objectives are as follows:

- Test and compare the performance of EfficientNetV2B2, Xception, ResNet50, VGG16, and InceptionV3, based on their accuracy and diagnostic effectiveness.
- Incorporate state-of-the-art Vision Transformer models, including MobileViT, Swin Transformer, DaViT, and MaxViT, to assess their potential in enhancing disease detection accuracy.
- Built a hybrid fusion model of Vision Transformers to leverage the strengths and achieve higher diagnostic performance.
- Analyze model outputs using Explainable AI techniques to provide transparency and insight into the decision-making process, making the diagnostic tools understandable and usable by non-expert users, such as farmers and agricultural stakeholders.
- Develop a reliable and lightweight diagnostic tool that can be implemented on various devices, ensuring practical usability for farmers, even in resource-constrained settings.

1.4 Contributions

This study's main contribution focused on building a model which is robust and applicable in real world application. In this regard, we conducted a thorough analysis of advanced state-of-the-art models that have demonstrated effectiveness in disease diagnosis. By comparing these trendy and well-performing models, we aim to create a solution that is both effective and applicable in real-world agricultural settings. We achieved it with the impressive performance of the fusion model. Additionally, To ensure transparency and build trust, especially for non-expert users like farmers, we incorporated various Explainable AI techniques. Another key contribution of this study is the successful implementation of the model on a Raspberry Pi hardware device. This demonstrates the model's practicality for real-world agricultural applications, ensuring that it can be deployed even in resource-constrained environments.

- The impressive performance of fusion model due to the combination of strengths from both VT models, MobileViT and DaViT.
- Successful implementation of Fusion model on Raspberry Pi.
- The enhancement on transparency by implementation of three XAI methods including LIME, Grad-CAM, and Saliency Map on the fusion model.

1.5 Structure of The Paper

The paper structure is illustrated in figure 1.2 as follows: Chapter One the introduction, aim, objectives, and problem statement. Chapter Two provides a summary of the literature review, discussing prior work in this field. Chapter Three details the methodology, including the dataset and models. Chapter four, Result and Discussion, presents the model performance and accuracy. Chapter five outlines the XAI and finally, Chapter Six discusses conclusion and future direction.

Together, the paper guides you step by step through the comprehensive process of developing and advanced maize disease diagnostic tool, from understanding the challenges to implementing and evaluating cutting-edge solutions. As we move forward, the goal is not just to advance technology, but to make a real difference in the world of agriculture and beyond.



Figure 1.2: Organization of the Paper.

Chapter 2

Related Work

Unlike other crops, maize, which is also popularly known as corn, is a versatile crop and can endeavor in different climates. Maize leaf abnormalities can be categorized, identified, and calculated using deep learning and machine learning techniques. This section assess previous studies on identifying corn leaf diseases.

In paper [43], explored the challenge of accurately identifying and categorizing diseases specifically Northern Leaf Blight of maize plant leaf. In addition to Gray Leaf Spot (GLS), and Northern Leaf Spot, in various environmental conditions. Using the CD&S dataset comprising 1,597 images, the study aimed to improve disease classification performance. It proposed MaizeNet, a deep learning model integrating Faster-RCNN with ResNet-50 and spatial-channel attention mechanisms. Through extensive experimentation, MaizeNet achieved a notable accuracy of 97.89%, demonstrating significant improvements in disease spot localization and overall detection accuracy. It effectively distinguished between a distinct class of corn leaf disease amidst cluttered backgrounds and lighting variations.

In the study [41], presented a novel mobile system used with regard to identification and the categorization of diseases in maize leaf, addressing the pressing issue of agricultural losses attributed to undetected or misclassified diseases, particularly prevalent in regions like Punjab, Pakistan. To confront this challenge, the researchers collected a diverse dataset comprising over two thousand images of maize leaf diseases in several growth stages, weather conditions, and time intervals. By employing deep learning some models used include from YOLOv3-tiny, up-to YOLOv8n, rigorous training and testing procedures were conducted, supplemented by meticulous image preprocessing and annotation techniques. Notably, YOLOv8n came up as the highly effective model, showing performance that is outstanding, with high precision and mAP (mean average precision) for disease detection and classification, achieving a commendable detection speed of 69.76 FPS. This research underscores the possibility of leveraging deep learning for real-time agricultural management of diseases, urging for proactive measures to minimize agricultural losses and boost crop yield. Moreover, the study suggests broader applications, advocating for integration with smartphone apps and UAVs to facilitate widespread adoption in agriculture, ultimately aiming to enhance global food security through sustainable crop management practices.

Similarly, another study [57], explored a comprehensive approach to boost the accuracy of corn leaf disease identification through the utilization of advanced technologies. By leveraging SVM (Support Vector Machine) alongside Convolutional Neural Networks such as AlexNet and ResNet50. The study aims to revolutionize disease identification in maize crops, crucial for global food security. Through the collection and preprocessing of a dataset comprising over three thousand corn leaf images from Embu County, Kenya, encompassing three disease categories, the researchers conducted rigorous experimentation and evaluation. The results demonstrate that CNN models, particularly AlexNet, outperformed traditional SVM classifiers, achieving remarkable rates of accuracy which are 98.3% and 96.6% respectively. The potentials of deep learning approaches is shown by this research in agricultural practices, supplying significant avenues in the enhancement of crop protection measures and contributing to sustainable agriculture worldwide.

In another work [13], the paper addresses the pressing issue of low agricultural productivity due to plant diseases, with a particular focus on maize plants. Emphasis was made on the significance of early detection of diseases in mitigating losses faced by farmers. Random Forest, Decision Tree, K-Nearest Neighbor, Naive Bayes, and Support Vector are all employed in supervised machine learning models. The study developed an accurate disease detection and classification models by analyzing high-resolution images of maize leaves and extracting relevant features. The Random Forest classifier proves itself as the highly reliable model, outperforming others by achieving a classification accuracy as high as 79.23%, underscoring its efficacy in identifying and categorizing the diseases of maize leaf. Utilizing a large dataset of a total of 3,823 images categorized into four groups: common rust, healthy, gray leaf spot, and northern leaf blight, the research shows an enhance agricultural practices and sustainability through early disease detection by leveraging machine learning algorithms.

Similarly, in [29], focusing on grapes and tomatoes, the VGG16 modeler was used to classify and diagnose leaf diseases in these crops. They utilize data augmentation techniques, hyperparameter tuning, and model optimization to enhance model performance. This study evaluates the model using different performance metrics, achieving an accuracy rate as high as 98.40% for grapes and 95.71% for tomatoes. Researchers point out a significant benefit of early disease diagnosis with regard to agriculture, as well as demonstrate the effectiveness of deep learning techniques aimed at enhancing crop management practices. Through hyperparameter tuning and model optimization, they demonstrate the potential of deep learning to revolutionize agricultural practices and increase food production.

Another work [28], Deep transfer learning, has been used to classify corn disease and healthy plants from leaf images. Using convolutional neural network (CNN) models and a collection of 3852 images, the study's researchers obtained an average prediction rate of 98.6%, which is a good performance. They used ten public CNN models through transfer learning and evaluated their performance using various metrics. The results emphasize the potential of deep learning to enhance agricultural practices by enabling rapid and accurate disease identification, thereby bringing precision to crop management and food production.

In the paper[26] a method for accurate detection of maize foliar disease in complex environments using LS-RCNN and CENet cascade network is proposed. LS-RCNN detects corn leaves, and CENet classifies them into four categories. This method uses a two-stage transfer learning strategy for better accuracy and faster training. The results demonstrate higher f1-scores and faster training than other methods. This paper presents the dataset with images of laboratory and natural environments, along with a discussion of data augmentation.

Similarly, another study [48] presents LeafDoc-Net, a strong and lightweight transfer-learning design for precisely identifying leaf diseases over numerous plant species, indeed with constrained image information. The approach combines DenseNet121 and MobileNetV2 models, upgrading them with consideration instruments, world-wide normal pooling layers, extra-dense layers with swish actuation, and batch normalization layers. Assessed on cassava and wheat leaf malady datasets, LeafDoc-Net beats existing models in most of the performance measurements, with potential for further enhancement and expansion in future research.

Existing deep learning methods for corn disease detection often prioritize accuracy over real-time performance, limiting their usefulness in practical settings. To address this, [36] propose a lightweight object detection algorithm based on an improved YOLOv5s model. Their approach incorporates a Faster-C3 module to reduce model complexity, while also enhancing the neck network with CoordConv and a modified CARAFE module to improve semantic information extraction and detection accuracy. Finally, they leverage channel-wise knowledge distillation during training to further enhance accuracy without increasing model size. This method achieves a good balance between accuracy and speed, which made it suitable for real-world corn disease detection applications.

In response to the global spread of maize diseases, a novel classification model using DenseNet201 and an optimized Support Vector Machine (SVM) has been developed to effectively identify maize leaf diseases. In the study [35], leverages the advanced image-classification capabilities of DenseNet201 and Bayesian optimization techniques to improve SVM performance, addressing challenges such as variable lighting and reflections in image analysis. The model was tested on a dataset of 4988 images, categorizing them into four classes: healthy, blight, common rust, and gray leaf spot. Impressively, the proposed model acquired an accuracy of 94.6 percent, significantly outperforming traditional SVM approaches, thereby enhancing agricultural productivity and disease management.

In the study [25], a deep learning approach named WG-MARNet was proposed for identifying maize leaf diseases, that addresses noise, background interference, and low accuracy. WG-MARNet utilizes wavelet threshold-guided bilateral filtering (WT-GBF) to reduce noise and decompose images for improved feature extraction. It then employs a multichannel ResNet architecture with an attenuation factor for optimized multiscale feature fusion and training stability. Finally, the model leverages PRelu and Adabound for enhanced convergence and accuracy. This approach achieved a promising average recognition accuracy of 97.96 percent and a detection time of 0.278 seconds per image, demonstrating its potential for precise maize dis-

ease control in fields.

Another recent study [44], underscores the significant impact of deep learning techniques in agriculture, particularly in weed, pest, and disease detection. The study focused on experimenting with different CNN architectures, including DenseNet201, MobileNet, VGG16, Hyperparameter Search, and InceptionV3. By fine-tuning these models on agricultural image data, it achieved excellent accuracy in detecting the disease. In particular, the DenseNet model with outstanding accuracy of 99.62%, MobileNet performed well with 91.85% accuracy, and VGG16 achieved 78.71% accuracy. Additionally, the study highlighted the data augmentation and feature fusion as critical steps in increasing the models' performance.

Likewise in the paper [21], a specialized model, MFaster R-CNN, was tailored for detecting corn leaf diseases based on Machine Vision detecting corn leaf diseases in agricultural environments. The model enhances the Faster R-CNN framework by incorporating a batch normalization processing layer and a mixed cost function to improve accuracy and convergence speed. The study used a dataset of 697 images showing different maize diseases taken in various weather conditions. Results showed that MFaster R-CNN performed better than other models in detecting these diseases. Showcasing its potential for practical applications in agricultural disease control.

Another study [24], introduces a smart way to detect diseases in maize leaves using a special computer model called MFF-CNN. This model is designed to tackle common challenges in disease detection, like changes in lighting, complex backgrounds, and unclear target areas. The MFF-CNN model outperforms other methods in detecting maize leaf diseases quickly and accurately. The study's experiments prove that the MFF-CNN model works well in spotting maize leaf diseases, even in tricky situations like overlapping areas and sparse targets. This method not only improves detection accuracy but also speeds up the process, making it a useful tool for diagnosing maize leaf diseases and potentially other plant diseases.

In the paper [45], explored the significant impact of biotic stresses, such as fungal, bacterial, and viral pathogens, on maize yield and emphasized the importance of identifying resistant genes to develop disease-resistant cultivars. Their study employs both machine learning and deep learning techniques to classify gene expressions in maize under normal and stress conditions. The machine learning algorithms used include Support Vector Machine, Naive Bayes, Decision Tree, K-Nearest Neighbor, and Ensemble, while a Bi-directional Long Short Term Memory (BiLSTM) network with a Recurrent Neural Network architecture is introduced for deeper gene classification. To boost algorithm feature selection, performance was conducted using the Relief feature selection algorithm. The findings highlighted the superior performance of BiLSTM compared with other algorithms. Crucially, several genes, including (S)-beta-macrocarpene synthase, zealexin A1 synthase, and others, were identified as differentially upregulated under biotic stress, marking them as key targets for enhancing maize resistance to pathogens.

Another study [31], involved the comprehensive analysis of Convolutional Neural

Networks such as MobileNetV2 and Xception modules for detection of plant disease. Among the CNN architecture employed, MobileNetV2 displayed great efficiency suitable for mobile devices while Xception being an extension of the Inception module has improved extraction capabilities as its feature. The research paper presents an ensemble module. This ensemble module combines the strengths of Xception and MobileNetV2 to improve the performance of plant disease detection. The ensemble approach is referred to as LEMOXINET. The ensemble model was able to achieve great results with 99.10% accuracy.

In the study [37], conducted research where they did a comprehensive and comparative analysis of the various deep-learning modules to predict cotton diseases. By utilizing fine-tuning Transfer Learning algorithms, the Xception module achieved the highest accuracy of 99.70% among all the modules used. The researchers selected the Xception module for their web-based application for Cotton disease prediction, which will assist farmers in early diagnosis of cotton disease, increasing cotton production.

In another paper [42], they studied methods of disease classification by using the triCNN architectures including Inception, Xception, and DenseNet169. The paper provides some overviews of the triCNN architectures with the aid of visual images. The paper presents some computerized methodologies for the detection of groundnut disease by using the ensemble method. To get an accurate disease prediction, the researchers used a fusion approach, i.e., combining the triCNN architectures. An accuracy of 98.46% performance was obtained when their proposed framework was applied to the groundnut leaf datasets.

In the paper [47] conducted research using machine learning-based automated disease detection to accurately detect disease. By combining EfficientNetB0 and MobileNetV2 on PlantVillage datasets with about 54,305 images, the accuracy of disease prediction was improved by 99.77%. This model shows a more dependable automated detection system for disease detection.

In another recent study [54] addresses the early identification and precise categorization of numerous diseases affecting maize plants, including corn smut, corn rust, corn leaf blight, corn mosaic virus, and corn stunt. The motivation for this study derives from the important need to reduce the suffering caused by these diseases, especially in humid, warm locations where maize is often farmed. The process involves creating a CNN-based model from a dataset including four types of maize diseases: rust, gray leaf spot, healthy, and leaf blight. The study discovered that the suggested deep learning model was highly accurate in diagnosing corn diseases, with the highest f1-score accuracy recorded at 99.83%. The dataset for the study was obtained from Kaggle and comprised around 8000 images separated into training, test, and validation data. The models' results showed that deep learning is good at precisely recognizing and distinguishing between different types of maize diseases, which has implications for increasing crop output and developing agricultural technologies. Future research might entail creating more advanced models for more accurate and efficient disease categorization in maize crops.

Similarly [49], addresses the difficulty of accurately classifying corn seed diseases utilizing advanced AI approaches, especially MobileNetV2 with feature augmentation and transfer learning. The motivation for this research stems from the growing relevance of precision agriculture and the necessity for precise assessments of agricultural goods such as corn seeds. The researchers hoped to increase the model's capacity to extract features and identify diseases by selectively picking MobileNetV2 and adding layers such as Average Pooling, Flatten, Dense, Dropout, and Softmax. The study used a comprehensive dataset of 21,662 maize seed images from a laboratory in Hyderabad, India, divided into four classes: broken, discolored, silk cut, and pure. The results showed that the suggested model obtained an accuracy of roughly 96% across all four classes, exceeding state-of-the-art models in terms of precision, recall, F1 score, and accuracy. Feature augmentation and transfer learning were critical in boosting the model's accuracy by reducing overfitting, accelerating training, and improving adaptability to various patterns in the data. This discovery has substantial significance for the agricultural business and farmers coping with maize seed diseases, as it provides a potential approach for improving precision agriculture and crop management. The researchers recommend investigating more model upgrades and applications in real-world agricultural settings to improve disease categorization accuracy and efficiency.

Likewise, The study [58] describes image recognition by identifying maize leaf diseases using a dataset created by Indian researchers, intending to provide excellent solutions for agriculture. The motivation originates from the desire to help agricultural researchers and farmers quickly diagnose and treat maize leaf diseases in order to increase crop output and quality. The methodology included preprocessing techniques such as image adjustment, normalization, and data enhancement with the resnet18 deep learning model, which is well-known for its performance in image identification tasks. The model performed well in disease categorization, with 98% accuracy, 95% precision, 95% recall, and 95% f1-score. The dataset included 2,341 images: 575 healthy corn leaves, 661 corn leaf spot images, 503 corn leaf rust images, and 602 corn leaf blight images. Future study seeks to improve the model's accuracy and generalization by evaluating more datasets, resulting in sophisticated image recognition solutions for agriculture.

In the paper [34], The authors utilize six CNN architectures, including Basic CNN, EfficientNetV2B0, EfficientNetV2B1, VGGNet, LeNet-5, and ResNet to detect the maize image diseases. The study use a dataset, consisting of 1,5344 maize leaf images, including MLN, MSV, and healthy. Moreover, to increase the performance of the models, the study performs hyperparameter tuning. The study results showed that the EfficientNetV2B0 model perform well with promising accuracy of 99.99% to detect the maize leave disease. Finally, the study implements Explainable AI method Grad-CAM for model interpretability.

In another research [53], focuses on the use of machine learning techniques, specifically Explainable AI and Deep Learning, to diagnose two prevalent diseases in maize crops: maize streak and maize leaf blight. The study utilize two pre-trained Transfer Learning models such as VGG19, and MobileNet for maize disease detection. Furthermore, the study implements two Explainable AI method shapley additive

explanations (SHAP), and local interpretable model-agnostic explanations (LIME) with best performing model for model interpretability.

In [52], the research examines the performance of six deep transfer learning pre-trained models such as VGG19, VGG16, MobileNetV2, ResNet50, ConvNextBase, and InceptionV3 to detect maize leaf diseases. The study use two maize leaves image datasets, namely PlantDoc, and Plant Village. Moreover, the research suggests a new method of combining the attention mechanism with ResNet50 and VGG16 models. The use of attention mechanisms has greatly enhanced the precision of detecting diseases in maize leaves. Overall, the composite VGG16+SE model achieve a validation accuracy of 93.44%, while the MobileNetV2 model excel with the best accuracy of 94.76% among all models.

In light of the challenges posed by disease vulnerability in the maize industry, the study [23], proposed an automated system for disease identification, severity assessment, and yield loss quantification in maize using a real-world dataset annotated by plant pathologists. The authors put forward a deep learning model called ‘MaizeNet’ that uses K-Means clustering for region of interest extraction and has a remarkable accuracy of 98%. 50% accuracy. The integration of the model into the ‘Maize-Disease-Detector’ web application provided a friendly user interface; thus, it is a valuable resource for plant pathology specialists. The high accuracy of the model, the ability to extract features, a small number of parameters, and the speed of training show the possibility of using the model to transform disease control in maize crops.

In the paper [38], the authors stressed on the significance of disease diagnosis at an early stage in crops to enhance the quality and quantity of crops. Conventional disease identification was a complicated process that called for expertise and time, which is why the authors proposed an automated system that would be very helpful in agriculture. They developed a stepwise disease detection model that involved images of diseased and healthy plants, with the images being passed through a CNN algorithm with five pre-trained models. This model was structured into three stages: crop classification, disease identification, and disease categorization with an ‘other’ category for increased model versatility. In validation tests, the model achieved a high level of accuracy in categorizing crops and disease types at 97. 09%. Further, the flexibility of the model was demonstrated when the accuracy of the model increased when non-model crops were included in the training data set. The study claimed that the model had a lot of potential for smart farming, especially for Solanaceae crops, and its applicability was believed to grow as more crop types were included in the training set.

In the study [39], discussed the limitations of automated crop disease detection, including data privacy and costs, in the context of federated learning. They experimented with CNN models, mainly ResNet50, and vision transformers (ViT) using a dataset from PlantVillage. Findings revealed that federated learning is efficient based on the number of learners and the quality of data. ResNet50 was the best suited to federated learning compared to ViTs because of the higher computational complexity. This study also highlighted the possibility of using federated learning

in crop disease classification and the directions for future research.

The paper [40] introduces an interpretable machine learning framework utilizing Convolutional Neural Networks (CNNs) and explainable AI (XAI) techniques to accurately diagnose Maize Streak Disease. The framework combines deep learning for precise disease classification with interpretability methods such as SHAP and LIME. The accuracy of 96% in identifying Maize Streak Disease, highlighting the effectiveness of the interpretable deep learning approach. The study emphasizes the importance of transparency and interpretability in deep learning models to enhance user trust and understanding in agricultural disease diagnosis.

Similarly, In [55], the study explores the application of deep learning and vision transformer models for detecting and classifying maize leaf diseases. The study evaluates the performance of various CNN architectures and vision transformers. CNNs, such as ResNet and DenseNet, have demonstrated high accuracy in disease detection, with reported accuracies between 94% and 99%. Vision transformers performed well in handling complex image data, offering detailed feature extraction and potentially superior performance.

In another comparative analysis paper, [50] the performance of different CNN models such as EfficientNetV2, MobileNetV3, ResNet50, and InceptionResNetV2 evaluated on image classification tasks. In the study, the InceptionResNetV2 achieved the highest accuracy of 88.9%, which is the combination of Inception and ResNet architectures. For the optimal performance, it requires fine-tuning. MobileNetV3 is optimized for low-latency applications and is ideal for mobile and edge devices with limited computational power. Also, ResNet50 offers good performance but demands significant computational resources, effective in training deep networks. The study concludes that the choice of CNN model depends on the specific requirements of the application, including computational resources and accuracy needs.

This study [11] presents a domain-specific vision dataset called DataDeep. The CropDeep is aimed at providing the data benchmark for a deep-learning-based classification and detection model construction based on realistic characteristics of agriculture. The CropDeep consists of 31, 1347 images with over 49,000 annotated instances from 31 different classes. These images were collected in a wide variety of situations using different cameras and greenhouse equipment. It also features visually similar species and periodic changes with more annotations, which have supported stronger benchmarks for deep-learning-based classification and detections. To verify the applications of the DeepCrop, deep learning models were performed on the data sets. In the process of ascertaining the applications of DeepCrop, a comparison of performances of seven deep learning models was used and classification and detection results were accumulated. VGG16, VGG19, SqueezeNet, InceptionV4, DenseNet121, ResNet18 and ResNet50. Out of these seven models, ResNet50 had the highest performance accuracy with 99.89% accuracy. In terms of detection results for Faster R-CNN, SSD, RFB, YOLOv2, YOLOv3, and RetNet. YOLOv3 obtained the second-highest average mAp of 91.44% and the study suggests that the YOLOv3 network has good potential in agriculture applications.

Another study [51] introduces an approach for Northern Leaf Blight detection as

early as four to five days using sensors of the Internet of Things (IoT). With the utilization of Convolutional Neural Networks and Long Short Memory(LSTM), ultrasound and Volatile Organic compound emissions were visualized and analyzed. A hybrid CNN-LSTM model was used to classify the Volatile Organic Compound, while an LSTM model was used for the classification of ultrasound detection from the maize crop. The hybrid CNN-LSTM model achieved a test accuracy of 96.39% after 15 training epochs in terms of Volatile Organic Compound classifications. A 99.98% accuracy was exhibited by the LSTM model in identifying anomalies in the ultrasound emissions from the maize plant.

Another study [20] performed deep learning approaches to identify maize disease. The dataset used for this study contains images of three diseases which include Maydis Leaf Blight, Turcicum Leaf Blight, and Banded Leaf and Sheath Blight. Utilizing the basic framework of the state-of-the-art InceptionV3 network, three network architectures were modeled on the dataset. The computational layers were trained with the dataset by the application of baseline learning. The Inception-V3_GAP was efficient in learning the features of the symptoms of the maize disease and thereby produced an accuracy of 95.99% in the separated dataset. To demonstrate the effectiveness of the proposed approach, a comparative analysis of pre-trained state-of-the-art networks was conducted. The results showed that the Inception-V3_GAP model involves higher computational cost. Besides the higher computational cost, the Inception-V3_GAP model performed quite better in terms of the classifications of diseases correctly based on the learned features from the dataset

Table 2.1: Summary Table For Selected Papers

Ref	Year	Proposed	Findings	Accuracy
[43]	2023	MaizeNet: Identification of Corn Leaf Illnesses using Deep Learning Methods	Showed significant improvements in disease spot localization and successfully distinguished various types of disease lesions amidst crowded backgrounds and lighting variations.	97.89%
[41]	2023	detecting and classifying maize leaf disease using deep learning and a mobile-based system	The research shows the effectiveness of YOLOv8n and the potential for real-time agricultural disease management.	N\A
[57]	2023	Employing Convolutional Neural Networks AlexNet and ResNet50 and Support Vector Machines to Identify Corn Leaf Diseases	They combined AlexNet and ResNet50 to identify maize leaf diseases accurately. The study shows that AlexNet outperform traditional SVM classifiers.	98.3%

Table continued from previous page

Ref	Year	Proposed	Findings	Accuracy
[13]	2020	Using Machine Learning Algorithms to corn Leaf Disease Detection and Classification	The study emphasizing the importance of early detection and potential of timely disease identification for farmers. They employed techniques like Naive Bayes and Random Forest.	79.23%
[29]	2021	Using VGG CNN for multi-crop leaf disease classification	Leaf diseases detection using deep learning. improve model performance by data augmentation and VGG model tuning.	98%
[28]	2022	Using deep transfer learning for maize diseases Classification	Utilized deep transfer learning for maize diseases classification. Explained deep learning’s potential in agriculture. Improves crop management and food production.	98%
[26]	2022	Maize Disease Identification using Cascade Networks & Two-Stage Transfer Learning	Introduces LS-RCNN and CENet for maize disease classification. Two-stage transfer learning boosts accuracy and training speed. Achieves high f1-scores. Includes dataset and discusses data augmentation.	99.70%
[48]	2024	Transfer learning-based architecture for accurate detection of leaf diseases in numerous plants using less amount of images	Surpasses current models in terms of AUC, recall, accuracy, and precision metrics on cassava and wheat leaf disease datasets. Emphasizes data augmentation and preprocessing, utilizes Grad-CAM++ for performance analysis, and shows promising results for generic leaf disease detection.	98%
[36]	2023	Efficient Model for Detecting Maize Leaf Disease using Knowledge Distillation	Improved the YOLOv5s model for detecting maize diseases by incorporating a Faster-C3 module, enhancing it with CoordConv and a revised CARAFE module, and utilizing channel-wise knowledge distillation.	mAP(0.5) accuracy.

Table continued from previous page

Ref	Year	Proposed	Findings	Accuracy
[35]	2023	Identification of maize diseases based on improved support vector machines using DenseNet201's deep features	Developed a classifier that incorporates DenseNet201 and SVM, improved with Bayesian optimization. This model effectively tackled imaging issues, such as lighting contrast changes.	94.6 %
[25]	2023	Maize leaf disease identification based on WG-MARNet	Used machine learning and deep learning techniques, including a Bi-directional Long Short Term Memory (BiLSTM) network, to identify maize genes that respond to biotic stress.	92.86%
[44]	2023	Crop Yield Improvement with Weeds, Pest and Disease Detection	The study highlighted the importance of data augmentation and feature fusion in getting better performance of each model. The models used in the study were the DenseNet, MobileNet, and VGG16.	DenseNet: 99.62%, MobileNet: 91.85%, and VGG16: 78.71%
[21]	2023	MFaster R-CNN for Maize Leaf Diseases Detection Based on Machine Vision	The specialized model introduced in the paper, MFaster R-CNN performed better than all other models in detecting diseases which has performed on a dataset containing 697 images.	97.18%
[24]	2022	One-Stage Disease Detection Method for Maize Leaf Based on Multi-Scale Feature Fusion	Comparative analysis of different CNN models, where MFF-CNN outperformed well even in handling overlapping and sparse targets. It can handle effectively challenges like changes in lighting, complex backgrounds, and unclear target areas that make it a feasible solution even for other plants disease detection.	N\A

Table continued from previous page

Ref	Year	Proposed	Findings	Accuracy
[45]	2023	Integrated transcriptomic meta-analysis and comparative artificial intelligence models in maize under biotic stress	Used a variety of ML and DL, including a BiLSTM network, to identify gene expressions in response to stress. BiLSTM demonstrated greater efficacy in identifying important genes such as (S)-beta-macrocarpene synthase, which are prospective targets for enhancing maize disease resistance.	92.86%
[31]	2022	LEMOXINET: Plant disease prediction using the Lite ensemble MobileNetV2 and Xception models	combine two CNN modules, MobileNetV2 and Xception to form an ensemble module called LEMOXINET with an accuracy of 99.10%.	99.10%
[37]	2023	A deep learning module for Cotton disease prediction using fine-tuning with a smart web application	Xception module is selected for the cotton disease prediction web application due to its high accuracy among all the Transfer Learning modules.	99.70%
[42]	2023	An ensemble of CNN models for detecting groundnut plant leaf diseases.	An accuracy of 98.46% was achieved from the combination of the tri-CNN architecture (Inception, Xception, and DenseNet169) in groundnut plant leaf disease detection.	98.46%
[47]	2023	Ensemble of deep learning models for multi-plant disease classification and smart farming	Combination of EfficientNetB0 and MobileNetV2 to improve plant disease classification accuracy.	99.77%
[54]	2024	Deep Learning for Classifying Corn Diseases	The study identified successfully maize diseases using deep learning, indicating its potential for enhancing crop productivity.	99.83%

Table continued from previous page

Ref	Year	Proposed	Findings	Accuracy
[49]	2024	Using MobileNetV2 with feature augmentation and transfer learning to enhanced corn seed disease classification	model for maize seed images. Feature augmentation and transfer learning boosted model accuracy, decreased overfitting, and accelerated training.	96%
[58]	2024	Using the ResNet18 model to identify maize leaf disease images	The model improved its disease classification performance by using approaches like image adjustment and data enhancement, surpassing other models using a dataset of 2,341 images.	95%
[34]	2024	Detecting Maize Lethal Necrosis and Maize Streak Virus using deep learning approach	Utilized six CNN architectures, including Basic CNN, EfficientNetV2B0, EfficientNetV2B1, VGGNet, LeNet-5, and ResNet to detect the maize image diseases.	99.99%
[53]	2023	XAI for Maize Disease detection	Utilized VGG19, and MobileNet to implement XAI for maize diseases detection.	N\A
[52]	2024	SE-VGG16 MaizeNet: Maize Disease Classification Using Deep Learning and Squeeze and Excitation Attention Networks	Examined the performance of six deep transfer learning pre-trained models such as VGG19, VGG16, MobileNetV2, ResNet50, ConvNextBase, and InceptionV3 to detect maize leaf diseases.	94.76%
[23]	2022	Disease detection, severity prediction, and crop loss estimation in MaizeCrop using deep learning	introduced ‘MaizeNet’, a deep learning model for maize disease management, achieving 98.50% accuracy and promising to revolutionize disease control through its web application integration and efficient training.	98%
[38]	2023	Developed a deep learning disease detection model for plants.	Implemented an automated, versatile CNN-based disease detection model for crops that proved highly accurate and showed potential for expanding smart farming practices.	97. 09%

Table continued from previous page

Ref	Year	Proposed	Findings	Accuracy
[39]	2023	Crop disease detection using images and federated learning.	Demonstrated that federated learning can effectively address data privacy and cost issues in automated crop disease detection, with ResNet50 outperforming other models.	ResNet50: 100%, ViT_B16 98.56%, Vgg16 & ViT_B32, 98.2%, InceptionV3, and 96.20%
[40]	2023	Interpretable deep learning for diagnosis of Maize streak disease	The study combines deep learning for precise disease classification with techniques for model interpretability, such as SHAP and LIME.	96%
[55]	2024	Maize Leaf Disease Detection Using Vision Transformers (ViTs) and CNN-Based Classifiers: Comparative Analysis	The evaluation performance of various CNN architectures and vision transformers. Vision transformers performed well in handling complex image data.	CNN(94% 99%)
[50]	2024	Classifying fine-grained maize leaf diseases using deep transfer learning	Among all the CNN models the combined architecture of Inception and ResNet has the highest accuracy. It balances accuracy and efficiency but requires fine-tuning for optimal performance.	88.9%
[11]	2019	CropDeep: The Crop Vision Dataset for Classification and Detection in Agriculture Using Deep Learning	CropDeep is a dataset aimed to provide a benchmark for deep-learning-based classification. Other models are also used to verify CropDeep Applications	For classification InceptionV4: 96.89% For Detection RetNet: 92.79%

Table continued from previous page

Ref	Year	Proposed	Findings	Accuracy
[51]	2024	Detecting non-visual maize disease using wave transform and hybrid CNN-LSTM models using VOC and ultrasonic IoT sensor data	With Convolution Neural Networks, LSTM, and IoT sensors, Northern Leaf Blight disease can be detected as early as for to five days after the occurrence of the disease.	Hybrid CNN-LSTM 96.39% LSTM 99.98%
[20]	2022	A deep learning-based method for identifying maize crop diseases	InceptionV3_GAP has a higher accuracy with higher computational cost but still performs better	95.99%

In summary, from the above discussion, it is distinctly noticed that most of the research in this field is classification and detection and most classification tasks are based on corn leaf disease classification and detection. However, there are some diseases which have not been analyzed. For instance, Maize Lethal Necrosis (MLN) disease classification, Moreover some research was focused on different crop disease detection as depicted in Table 2.3. Furthermore, Minor works on maize leaf disease detection using deep learning has been conducted, maize disease detection using explainable artificial intelligence should be more prominent. In this digital era agriculture should not be left behind with the use of technology, therefore creation of a detection tool for farmers in Africa and the rest of the world especially in Tanzania who mostly face those diseases is essential.

Table 2.3: Comparison of Different Papers With Our Paper

Paper	Year	Maize Disease Analysis	MLN	MSV	Dataset used	ML & DL	XAI
[48]	2023	-	-	-	-	≡	≡
[57]	2024	≡	≡	≡	=	≡	-
[29]	2022	-	-	-	≡	≡	-
[21]	2022	≡	-	-	-	≡	-
[28], [26]	2022	≡	-	-	=	≡	-
[43], [41], [13]	2023	≡	-	-	=	≡	-
Our paper	2024	≡	≡	≡	≡	≡	≡

≡ Covered = Partially Covered - Not Covered

Dataset Image used < 1000: -

Dataset image used 1000 to 5000: =

Dataset Image used > 5000: ≡

Chapter 3

Methodology

3.1 Work Plan

The work plan for this thesis followed a structured and systematic approach to develop and evaluate advanced deep learning models to detect maize diseases as it can be seen in figure 3.1. The research was divided into key stages to ensure a smooth flow of work from data collection to model evaluation and deployment.

Initially, the dataset was gathered over a six-month period using the AdSurv mobile application. This was followed by a critical preprocessing phase, where the images were cleaned, resized, and normalized to meet the input requirements of various models. Next, the data was used to train and assess several Transfer Learning and Vision Transformer models. These models incorporated EfficientNetV2B2, ResNet50, InceptionV3, VGG16, Xception, as well as Vision Transformers like Swin, DaViT, MobileViT, MaxViT, and the Involutional Neural Network (INN). During this phase, hyperparameters were fine-tuned to improve model performance.

Following the training and evaluation of individual models, a fusion model combining the strengths of MobileViT and DaViT was developed. This hybrid model achieved outstanding results and was further tested for its robustness and accuracy across various test scenarios.

The final stage involved deploying the fusion model on a Raspberry Pi, demonstrating the practical applicability of the solution to be effective in resource-constrained environments. Additionally, XAI techniques such as Grad-CAM, LIME, and Saliency Maps were incorporated to ensure model transparency and usability for non-expert users, like farmers.

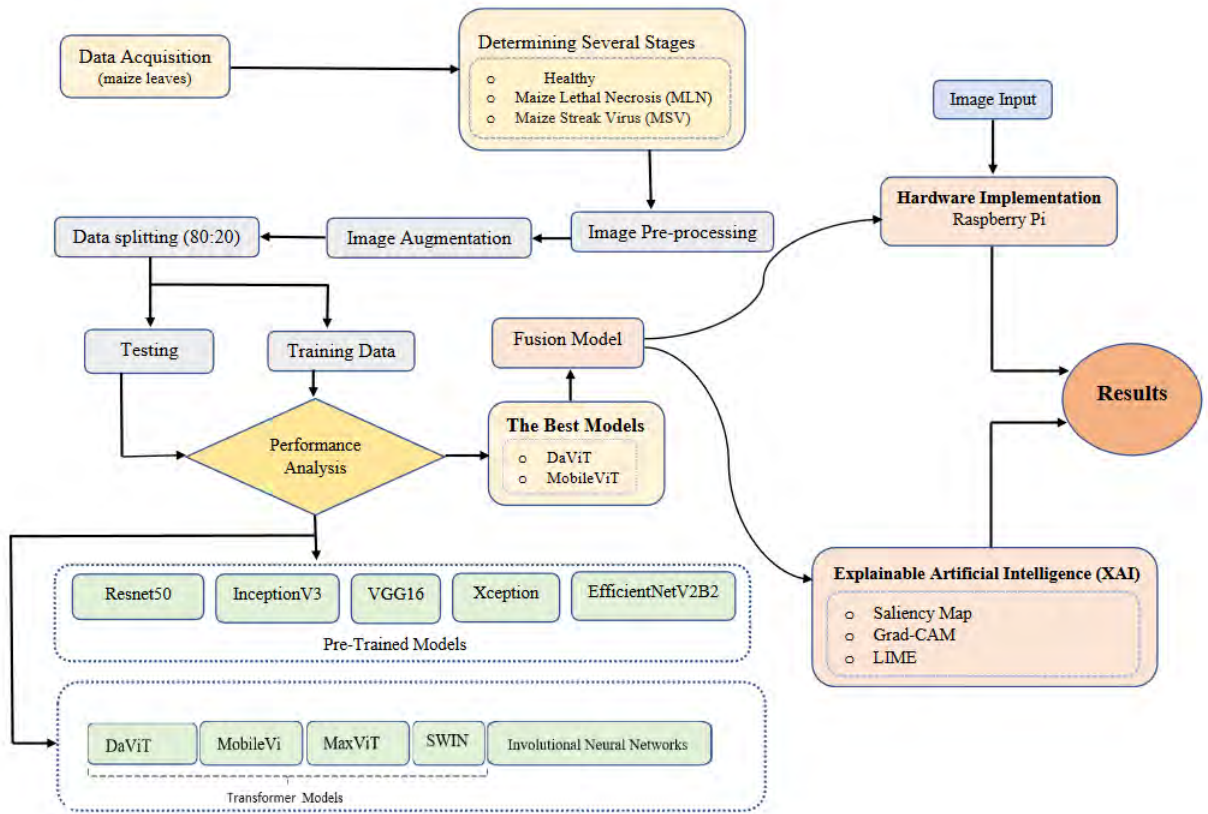


Figure 3.1: Diagram of the Work Plan.

3.1.1 Data Collection

The data collection process involved acquiring maize leaf images from farmers' gardens in Tanzania using the AdSurv mobile application installed on Samsung phones. A group of researchers and students from Tanzania Agricultural Research Institute and The Nelson Mandela African Institution of Science and Technology collected the dataset for a period of six months, from February 2021 to July 2021. The images were collected to diagnose MLN and MSV diseases as shown in figure 3.2, aiming to assist farmers in disease diagnosis and improve maize production.

The dataset consists of 17,277 labeled images categorized into Healthy (5,542), Maize Lethal Necrosis (5,068), and Maize Streak Virus (6,667) as depicted by figure 3.3. Each image instance includes the crop status, variety, age, and location (district, sub-county). The data collected is well-labeled and curated, providing an open and accessible maize image dataset for machine learning experiments.

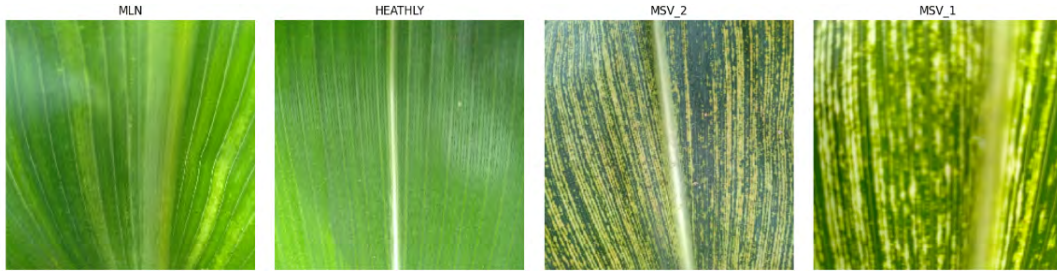


Figure 3.2: Sample images.

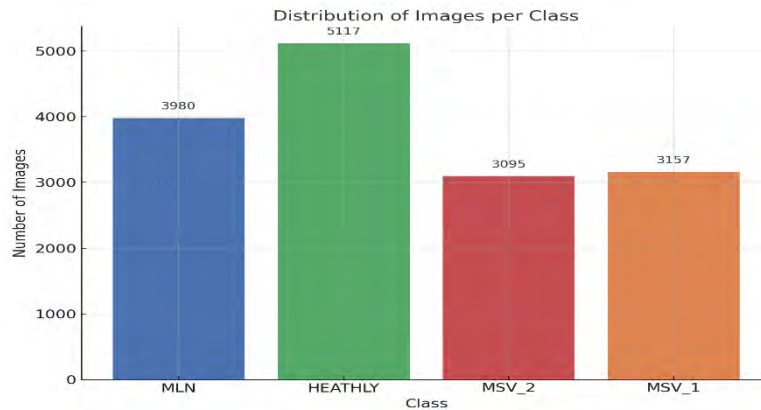


Figure 3.3: Distribution of Images in Each Class.

3.1.2 Data Preprocessing

For this thesis, the preprocessing of the maize disease detection dataset played a crucial role in ensuring the accuracy and robustness of the models. The dataset, containing images of maize leaves classified into Healthy, MLN, and MSV, underwent a series of essential preprocessing steps.

Firstly, all images were scaled to meet the input size requirements of each model. For instance, MobileViT required images to be resized to 224x224 pixels, while Swin Transformer worked with 256x256 pixel images. After resizing, normalization was applied to ensure that the pixel values were standardized. The normalization process used the mean and standard deviation values computed from the training dataset, helping the models to learn consistently across all image inputs. To enhance model generalization and improve performance, data augmentation techniques were employed. These techniques included random horizontal flips, rotations, zooms, and cropping, introducing variability in the training data. By applying these augmentations, the models became more resilient to variations in the maize images, such as lighting conditions and leaf orientations, which are common challenges in real-world scenarios.

Considering that the dataset is rather unbalanced, with certain classifications, like MSV, had more instances than others, a careful approach was taken to ensure that the models learned equally well across all categories. Weighted cross-entropy loss functions were implemented during training, compensating for the class imbalance and ensuring that predictions were not skewed toward more frequent categories.

The dataset was also split into training, validation, and test sets with an 80-10-10 ratio. The training set was used to train the models, while the validation set helped monitor performance and adjust model parameters. The test set, which was kept entirely separate, was reserved for evaluating the final performance of the models.

Efficient data loading was another critical part of the preprocessing pipeline. The dataset was batched, with batch sizes ranging between 16 and 32 images, depending on the available GPU resources. On-the-fly processing, including augmentation and normalization, was incorporated into the data pipeline, ensuring that the images were preprocessed in real time during model training. This approach minimized memory bottlenecks and sped up training times. Together, these preprocessing steps, resizing, normalization, augmentation, class balancing, and efficient batching were crucial in preparing the dataset for the deep learning models used in this research. They enabled the models to achieve high accuracy in maize disease identification, leading to reliable and practical outcomes.

3.2 Transfer Learning Models

Using a pre-trained model to increase learning efficiency on new tasks is known as transfer learning, a machine learning technique. It improves performance by transferring information from a source domain to a target domain, especially when the training data is inadequate or out-of-date. In computer vision, this technique has shown to be rather successful, particularly for applications like diagnosis and prediction[22]. Since AlexNet’s victory in the ImageNet competition, convolutional neural networks (CNNs) have been crucial to many deep learning tasks, frequently employing transfer learning. This method involves adapting a model trained on a large dataset to a related, smaller task. For example, a model trained on a large image classification dataset can be fine-tuned to categorize specific categories such as dogs and cats. The model’s learned features, such as edge or pattern detection, are either reused or refined to enhance performance on the new task[27]. Unlike multitask learning, which learns numerous tasks at once, transfer learning focuses on gradually transferring knowledge, making it excellent for settings that require progressive training and adaptability.

3.2.1 ResNet50

To improve the capacity to train deep networks, adding more convolutional layers by using residual learning is achieved using skip connections which successfully addresses the vanishing gradient problem figure 3.5 a transfer learning ResNet50 model was constructed[17]. It is a pioneering deep convolutional neural network built by Microsoft Research in 2015. It has 50-layer design where its architecture as depicted in figure 3.4 is separated into four major components: convolutional layers for feature extraction, identification blocks, convolutional blocks for feature modification, and fully connected layers for classification[7]. Furthermore, It was trained on the large-scale ImageNet dataset achieved a remarkable top-5 error rate of 6.71%, which is comparable to human performance. Moreover, It is the favored model for a many image classification applications, including medical image analysis, object identification, and facial recognition, because of its high accuracy, rapid convergence and

quick training [4].

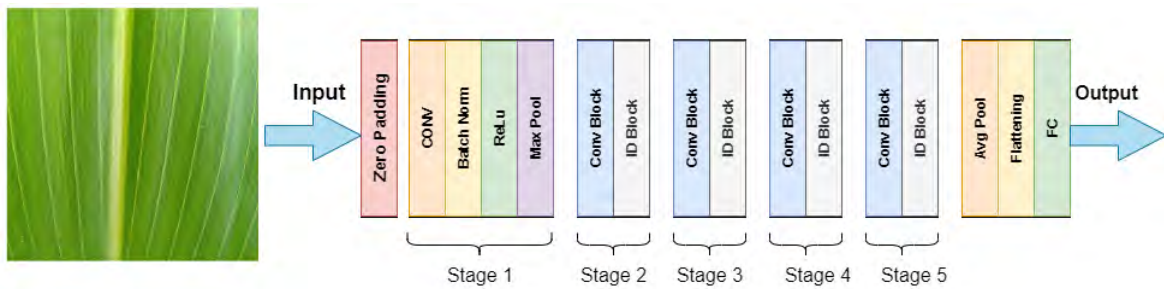


Figure 3.4: ResNet50 Model Architecture.

How ResNet50 solved the disappearing gradients' problem:

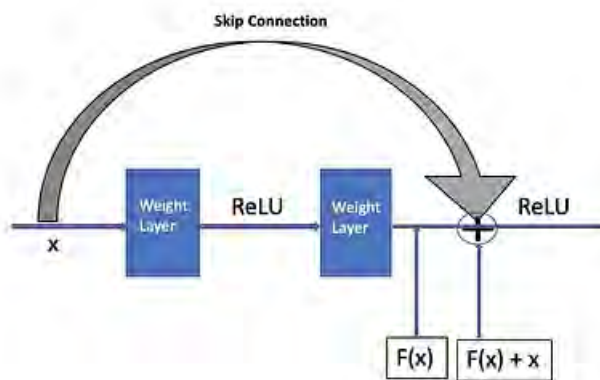


Figure 3.5: Skip Connection.

3.2.2 InceptionV3

InceptionV3 model is used in image identification. This model is a convolutional neural network architecture created by Google researchers in 2015, marks a vast step forward in computer vision. It has been built on the original Inception designs V1 and V2, it is intended to be computationally efficient while maintaining outstanding performance in image categorization applications. To extract features from images, the architecture employs a succession of convolutional, pooling, and inception modules. Furthermore, Inception modules enable the network to learn features at various scales and resolutions by performing numerous simultaneous convolutional operations of varying sizes. Moreover, it has showed world-class performance in a variety of computer vision tasks, including object identification, image classification, and visual question answering. It attained a 21.2 percent top-1 error rate and a 5.6 percent top-5 error rate in the 2012 ImageNet Large Scale Visual Recognition Challenge for single-frame evaluations [5]. Therefore, these performance measurements highlight InceptionV3's remarkable accuracy and efficiency, making it the preferred choice for difficult computer vision applications and cementing its status as the top deep learning architecture. In addition, InceptionV3 obtained a performance of 80%

accuracy, an 75 percent f1-score, and a recall of 76 percent in Maize disease identification using an 80:20 training-to-testing dataset. This demonstrates InceptionV3’s effectiveness in image classification and detection applications.

3.2.3 VGG16

This is a kind of artificial neural network introduced by K. Simonyan and A. Zisserman of the University of Oxford, It has become a key in the field of computer vision since its release in 2014. This model, which finished second in the ILSVRC 2014 classification challenge [2], is known for its basic yet successful design of 16 layers, comprising convolutional layers with modest 3x3 filters, max-pooling layers, and fully linked layers as shown in figure 3.6 [8]. Furthermore, it achieved an outstanding 92.7 percent top-5 test accuracy on the ImageNet dataset, which comprises over 14 million images from 1000 classes [3]. Moreover, by substituting bigger kernel-sized filters with many 3x3 filters, it improves on previous models such as AlexNet, allowing for deeper networks with more parameters. In addition, its design is distinguished by a constant input size of 224x224 RGB images, consistent usage of rectified linear units (ReLU), and the lack of Local Response Normalization (LRN), which reduces computation time and memory consumption. VGG-16 was trained on NVIDIA Titan Black GPUs, which is still an effective technique for large-scale image recognition.

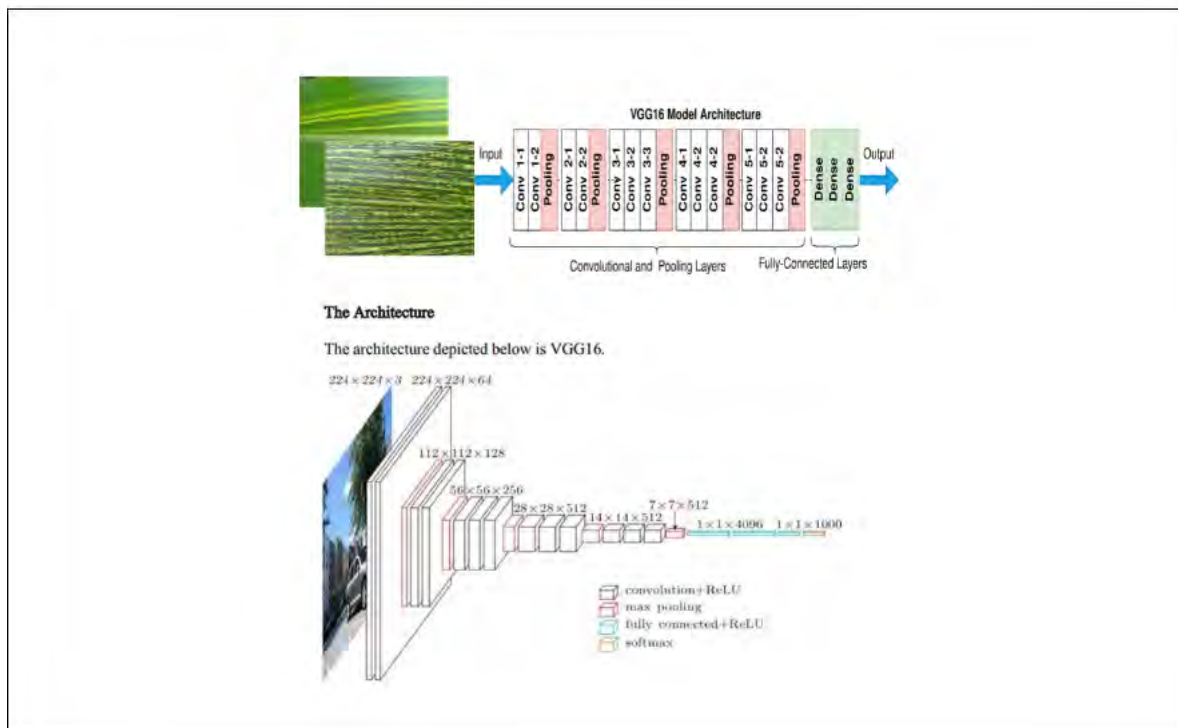


Figure 3.6: VGG16 Architecture.

3.2.4 EfficientNetV2B2

The EfficientNetV2 model developed by Mingxing Tan and Quoc V. Le is an innovative convolutional neural networks that achieve higher quicker training speeds and parameter efficiency than earlier models [10]. It was created through training-aware neural architecture search and scaling, improves both model size and training speed while including new procedures like Fused-MBConv as depicted in figure 3.7 [16]. This method allows EfficientNetV2 to be up to 6.8 times smaller and much quicker than other models. Furthermore, the design enhances further progressive learning by adaptive increasing regularization in parallel with image size, ensuring accuracy while preventing overfitting [16]. Moreover, it surpassed most current Vision Transformer (ViT) by 2.0% in accuracy and trained 5x-11x quicker with the same computing resources.

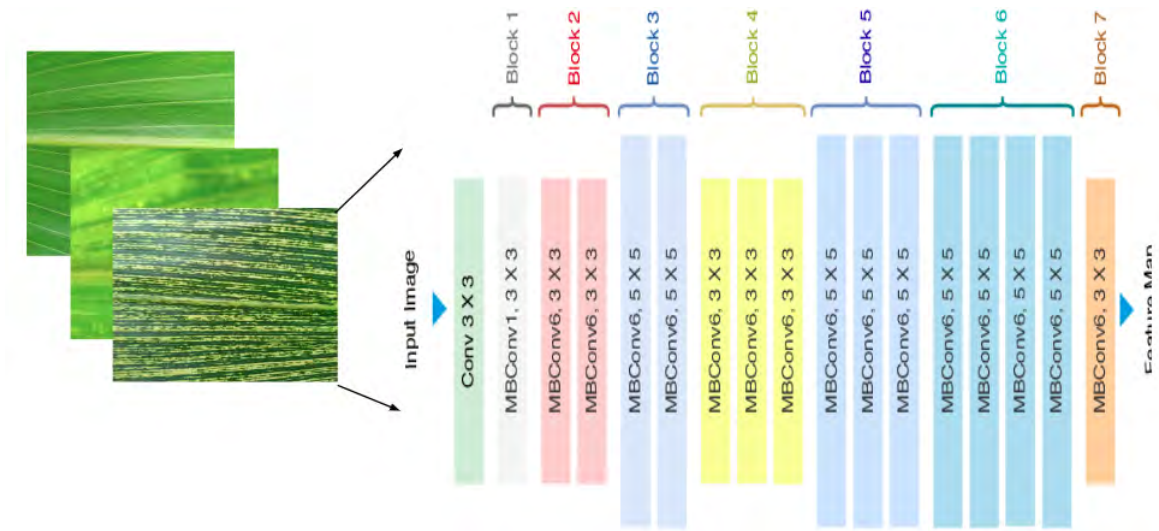


Figure 3.7: EfficientNetV2 Structures.

3.2.5 Xception

Xception model created by François Chollet in 2017 introduces an important step forward in convolutional neural network (CNN) design by using depth-wise separable convolutions, which separate spatial and depth operations to reduce parameters and computational costs while maintaining high computational power. This method enables Xception to surpass InceptionV3, particularly for large-scale image classification tasks. Xception outperforms InceptionV3 on both ImageNet dataset and on a larger dataset with 350 million images with 17,000 classes, all without increasing the number of parameters [6]. Furthermore, Xception's architecture is made up of entry and exit flows, which are strengthened by ResNet-inspired skip connections, and it uses global depthwise separable convolutions in its final layers to record global context. Additional tactics like data augmentation and batch normalization help to ensure quick training and higher results. Moreover, in our study we employed the Xception Model in Maize disease classification and achieved a performance of 89% accuracy, 86% f1-score, and a recall of 86% utilizing 80:20 training to test data with 10 input data epochs. Therefore, Xception delivers outstanding results, establishing

it as a robust and efficient model for a variety of computer vision tasks.

3.3 Vision Transformer Models and Hybrid Model

The Vision Transformer (ViT) is a unique architecture that uses the Transformer model, which was initially created for natural language processing applications, to do image identification at scale. It divides an image into sequences of flattened 2D patches, each handled as a "token," similar to how words are processed in NLP tasks, and then feeds them into the Transformer encoder. Unlike typical convolutional neural networks (CNNs), which use convolutional layers to identify local characteristics and generate hierarchical representations, ViT uses self-attention techniques to capture global dependencies in the image from the start [12]. This method reduces the requirement for handmade architectural components tailored to images, providing a more adaptable and scalable solution for image categorization problems. One of the primary benefits of applying a pure transformer model directly to image patch sequences is its ability to model global context without being constrained by the locality of convolution operations, potentially leading to improved performance on tasks requiring a holistic understanding of the image as depicted in fig 3.8. Pre-training ViT on large datasets, such as JFT-300M, and using transfer learning significantly improves its performance on various benchmarks by providing rich feature representations and improving generalization, outperforming state-of-the-art results on datasets such as ImageNet, CIFAR-100, and VTAB. ViT's efficiency in reaching competitive accuracy with fewer processing resources makes it an appealing alternative for picture classification tasks, demonstrating its potential to transform computer vision applications. To encode an image, we break it into fixed-size patches, linearly embed each one, add position embeddings, and use a typical Transformer encoder [9]. To classify a sequence, we often include a "classification token" that may be learned.

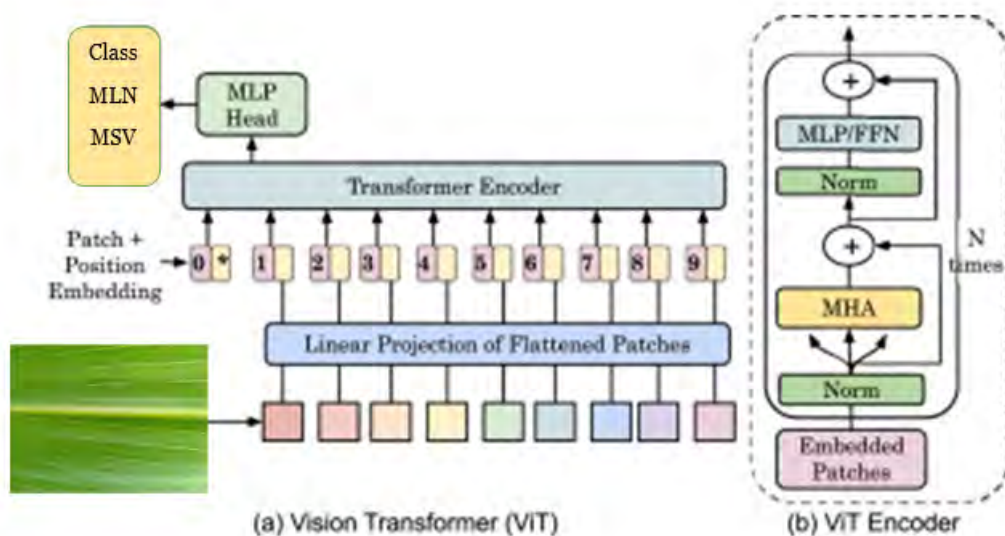


Figure 3.8: ViT Structures.

In our study, we used Vision Transformer (ViT) models such as SWIN ViT, DaViT, MaxViT, and MobileViT, which use self-attention processes to capture global visual features. These models are appropriate for tasks that need fine-grained feature extraction and generalization across many visual datasets. Furthermore, we used a hybrid model called the Involution Neural Network, which combines CNN efficiency with improved spatial feature learning via involution operations. These models were chosen based on their proven accuracy and adaptability in computer vision tasks, particularly when dealing with complicated patterns and little data.

3.3.1 Shifted Window Transformer (SWIN)

The Swin Transformer is a unique vision Transformer architecture intended to act as a flexible backbone for a variety of computer vision tasks. It proposes a hierarchical method to visual data processing using Shifted Windows, which enables fast calculation of self-attention inside non-overlapping local windows. This design option addresses fundamental differences between the language and vision domains, such as visual entity scale and image pixel resolution. The Swin Transformer’s hierarchical architecture has substantial advantages for modeling at many sizes, allowing the model to capture characteristics at numerous levels of abstraction. The Swin Transformer creates hierarchical feature maps by merging nearby patches in deeper layers, which may then be used with advanced approaches such as feature pyramid networks and U-Net [14]. One of the Swin Transformer’s important breakthroughs is its shifted windowing method, which enhances computational efficiency in self-attention computing by lowering the number of tokens required to interact while still maintaining the connection between neighboring windows. This method greatly decreases processing complexity when compared to standard Transformers, making it more efficient for high-resolution imagery. The shifted windowing technique increases computational efficiency by confining self-attention computation to local windows, but it also allows for cross-window connections, which enhances the model’s capacity to capture long-range relationships. As shown in figure 3.9. By combining grayscale image patches in deeper levels, the Swin Transformer generates hierarchical feature maps. Because each local window (red) uses self-attention processing, the computation cost is linear with the size of the input image. For applications involving dense recognition and picture categorization, it might serve as a general-purpose backbone. Conversely, prior vision Transformers [12] had a quadratic computing cost in respect to input image size because of global self attention computation, and they produce feature maps with a single low resolution.

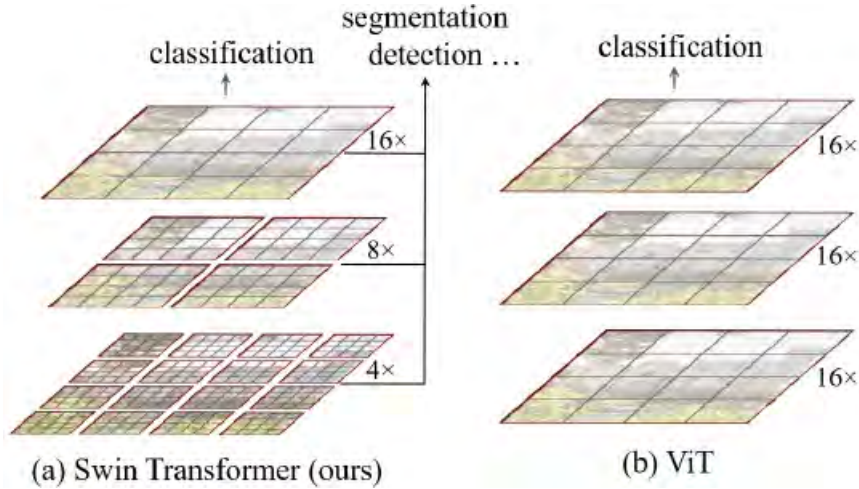


Figure 3.9: Swin Transformer vs ViT

In the context of image classification tasks, the Swin Transformer’s linear computational complexity with respect to image size, together with its good performance on benchmarks like ImageNet-1K, make it an appealing option for jobs that need accurate and quick visual modeling. Its hierarchical and efficient architecture makes it ideal for processing high-resolution pictures and complicated visual data, resulting in improved accuracy and performance.

3.3.2 Dual-Attention Vision Transformer (DaViT)

The model called the DaViT (Dual Attention Vision Transformer) model was presented to improve vision by including both spatial and channel attention. The model is built on a dual attention strategy, in which the channel attention model focuses on dependencies across feature channels, and the spatial attention model focuses on the interaction between various spatial positions of an image. This allows complex visual patterns managed by DaViT with improved capabilities and delivers better feature representations. It consists of several transformer layers, each with an incorporated spatial and channel attention mechanism. While the spatial attention allows the model to focus more on those crucial areas of input by re-weighting spatial locations, the channel attention amplifies features based on the relevance of specific channels to capture more contextual information in greater detail [18]. Some of the key features of DaViT are the Dual Attention Mechanism, which integrates both spatial and channel attention within a single framework for the exploitation of image spatial relations and channel dependencies. Multi-Scale Feature Extraction also combines different-scale features to enable the model to process variously-sized objects, enhancing its generalization capability. Another feature is the efficient computing system which balances high performance and computational cost. It can be applied to a range of vision applications, including segmentation, object identification, and image classification. DaViT is a more accurate and effective solution than traditional vision transformers for today’s visual issues.



Figure 3.10: DaViT Architecture

In figure 3.10, the DaViT architecture has its focus on the dual attention mechanism. It is a sequence of transformer blocks with spatial attention and channel attention as two important components. Spatial attention shall be used for refinement in the relationship between the features in space, while the channel attention shall enhance the dependencies along with inter-channels. These modules work together through multi-layers to process the input images, extract multi-scale features, combine spatial and channel attention, and promote feature representations. It thus can let DaViT handle complex visual patterns efficiently, which is very useful in tasks such as image classification and segmentation.

3.3.3 MobileViT

MobileViT is a powerful deep learning architecture designed to combine the strengths of CNNs (Convolutional Neural Networks) and transformers, making it suitable for mobile devices and resource-limited environments. MobileViT blends the strengths of CNNs and transformers, each playing a pivotal role in enhancing model performance. Where CNNs excel at capturing local features, like edges and textures, much like how our eyes first recognize the details in an image. Meanwhile, transformers are great in capturing and understanding long-range dependencies which provide a broader context that helps connect these finer details to the bigger picture. By combining both approaches, MobileViT ensures the model can grasp both intricate local patterns and overarching global relationships that result in stronger performance across tasks like image classification and detection [15]. One of the standout features of MobileViT is its lightweight and efficient design, specifically created to run on devices with limited computational power, like smartphones. Convolutions and transformers are used by MobileViT such that the resulting MobileViT block exhibits convolution-like characteristics and permits global processing, as shown in figure 3.11. This modeling capability allows us to design shallow and narrow MobileViT models, which in turn are light-weight. While transformers are typically resource-intensive, MobileViT scales them down to maintain high accuracy without sacrificing speed or memory efficiency.

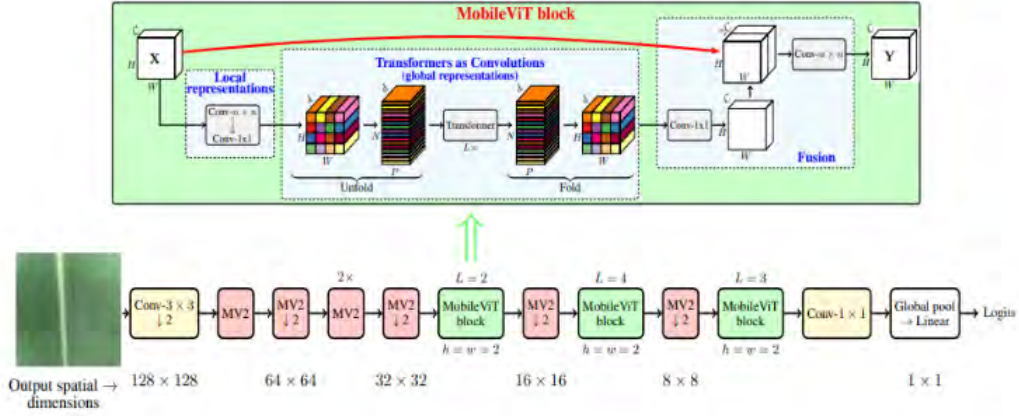


Figure 3.11: MobileViT Architecture

MobileViT Architecture: Here, in this case, MV2 stands for MobileNetv2 block, while Conv- $n \times n$ in the MobileViT block denotes a typical $n \times n$ convolution. Blocks with $\downarrow 2$ are those that use downsampling.

3.3.4 MaxViT

The architecture known as MaxViT which stands for Maximum ViT was first introduced in [32] and attempts to combine a convolutional neural network (CNN) and transformers to perform better on tasks by capturing both local and global features in images. All these achieve better performance in different vision tasks. This model utilizes multi-axis attention in improving the understanding of images. With images split into multiple patches, MaxViT is able to capture dependency in multiple directions, hence paying attention to local details and global information. Moreover, its dynamic attention layers improve classification and detection tasks by increasingly focusing on more informative parts of the image. With flexibility in scaling depth and width, MaxViT effectively balances computational efficiency and representation power, achieving state-of-the-art results on numerous vision benchmarks.

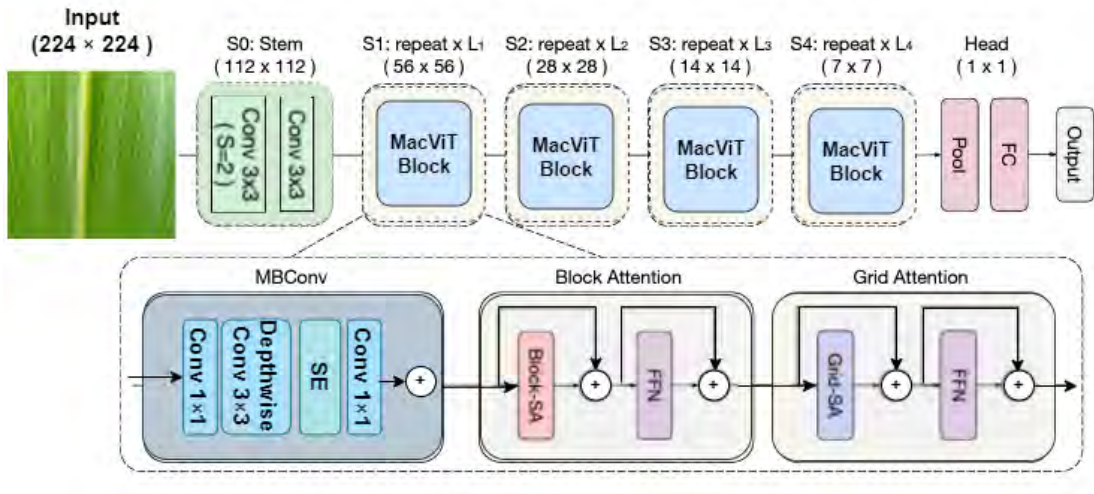


Figure 3.12: MaxViT Architecture

As shown in figure 3.12 We follow a typical hierarchical design of ConvNet practices (e.g., ResNet) but instead build a new type of basic building block that unifies MBConv, block, and grid attention layers. Normalization and activation layers are omitted for simplicity.

3.3.5 Involutional Neural Network

Involutional Neural Networks (INNs) are a recent architectural paradigm designed to replace standard convolution operations with a more efficient and adaptable process known as involution. Unlike convolution layers, which apply a fixed spatial kernel uniformly across all channels and pixels, involution utilizes dynamic kernels that are generated locally for each spatial position. This allows the model to flexibly adapt its operations to each pixel, improving both efficiency and context-awareness—particularly valuable when handling large spatial data, such as plant images. In this maize disease detection thesis, the use of INNs helps capture localized disease patterns more effectively by enabling the model to dynamically adjust its focus based on the surrounding context. This makes INNs more robust and efficient compared to traditional convolutional models.

Beyond that, involution introduces a novel approach that is both location-specific and channel-agnostic. Traditional convolutions face limitations in processing variable-resolution input tensors due to the fixed nature of their kernels. Involutions solve this issue by generating each kernel conditioned on specific spatial positions, as shown in the accompanying diagram. This allows the model to process input data at different resolutions with ease, improving adaptability to local variations in the input images. Based on the idea of maize disease detection, this dynamic kernel generation enhances the model’s ability to detect fine-grained disease symptoms, making it a valuable tool for recognizing intricate disease patterns and ensuring accurate diagnosis.

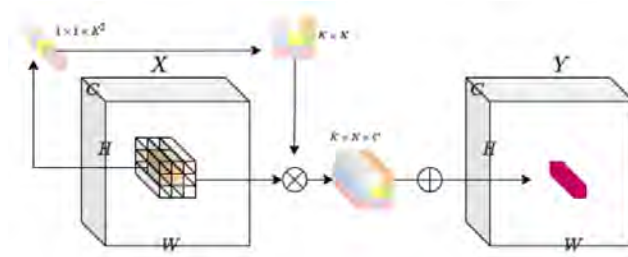


Figure 3.13: INNs Architecture

3.4 Proposed Fusion Model

To optimize the performance of our maize disease classification model, we created a fusion model by combining our trials’ top-performing Vision Transformers: MobileViT and DaViT. Both pre-trained and fine-tuned models on our dataset demonstrated high accuracy in feature extraction and classification. MobileViT was chosen for its lightweight architecture and computational efficiency, making it appropriate for real-time applications, as well as its use of convolutional layers and self-attention techniques to collect both local and global information. DaViT, with its dual-attention mechanism, excelled at handling complicated patterns, providing richer feature representations, particularly in demanding settings such as variable lighting. Our fusion method blended both models’ final layer outputs, resulting in a unitary feature vector that incorporated MobileViT’s local details with DaViT’s larger contextual knowledge. This vector was then fed through fully connected layers for categorization, allowing the model to make more accurate predictions as depicted in figure 3.14. The fusion model beat each individual model in terms of accuracy, generalization across a wide range of situations, and robustness, especially when one model struggled with unique data changes. This method demonstrates how merging transformer-based models can boost performance and efficiency in complex image classification problems.

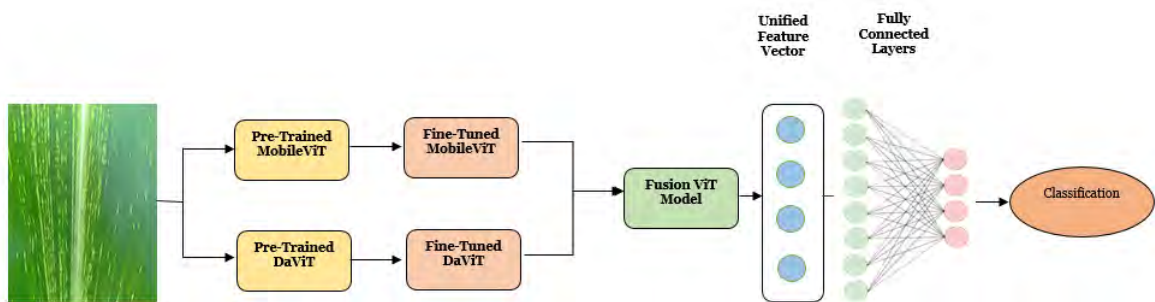


Figure 3.14: Fusion Model Structure

Chapter 4

Results and Discussion

4.1 Transfer Learning Model’s Performance

A comprehensive study was undertaken to improve the effectiveness of models., as shown in table 4.1. To ensure uniformity, many CNN architectures were explored utilizing standardized input sizes of 224x224x3. ResNet50, using a batch of size 32, Adam optimizer, learning rate of 0.001, and Categorical Crossentropy loss, used residual connections to decrease overfitting while avoiding dropouts. InceptionV3 and VGG16 employed the same batch size and optimizer, but with a 0.0001 learning rate and a 0.5 dropout rate to prevent overfitting. Xception used a learning rate of 0.001 and 0.5 dropout for depthwise separable convolutions. EfficientNetV2B2, which uses a dynamic learning rate, used built-in regularization rather than dropout to ensure a balance between learning capacity and model generalization.

Table 4.1: Parameter Settings for Different Models

Parameter	ResNet50	InceptionV3	VGG16	Xception	EfficientNetV2B2
Batch Size	32	32	32	32	32
Optimizer	Adam	Adam	Adam	Adam	Adam
Learning Rate	0.001	0.0001	0.0001	0.001	0
Input Size	224×224×3	224×224×3	224×224×3	224×224×3	224×224×3
Dropout	0	0.5	0.5	0.5	0
Loss Function	Categorical Crossentropy				

4.1.1 ResNet50

Training the datasets with the ResNet50 Model, 10 epochs, 383 batches and a data split ratio of 2:8 was used, with 80% for training and 20% for validation which yielded an accuracy of 91%. To aid understanding, the following confusion matrix in figure 4.1 and table 4.2 highlight the key trends and patterns observed in training the dataset with the ResNet50 model

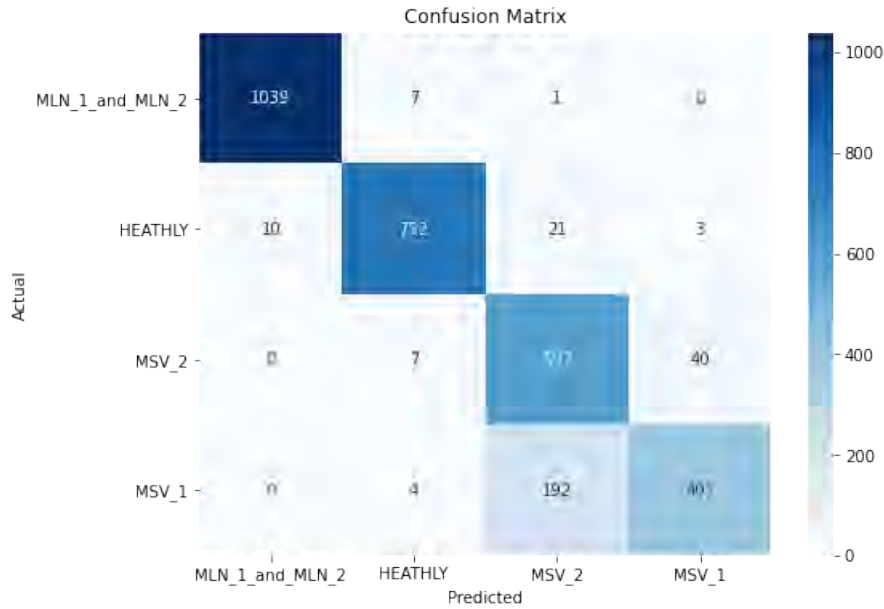


Figure 4.1: ResNet50 Model Confusion Matrix

Class	precision	recall	f1-score	support
HEALTHY	0.99	0.99	0.99	1023
MLN_1_and_MLN_2	0.98	0.96	0.97	796
MSV_2	0.73	0.93	0.82	631
MSV_1	0.90	0.67	0.77	619
Accuracy			0.91	3069
macro avg	0.90	0.89	0.89	3069
weighted avg	0.92	0.91	0.91	3069

Table 4.2: ResNet50 Model Report.

4.1.2 InceptionV3

Employing the InceptionV3 model to the dataset with 10 epochs, 383 batches and a data split ratio of 2:8 was used, with 80% for training and 20% for validation which yielded an accuracy of 91%. To visualize the findings, the following training and validation accuracy & loss in figure 4.12 and table 4.10 highlight the key trends and patterns observed in training the dataset with the InceptionV3 model.

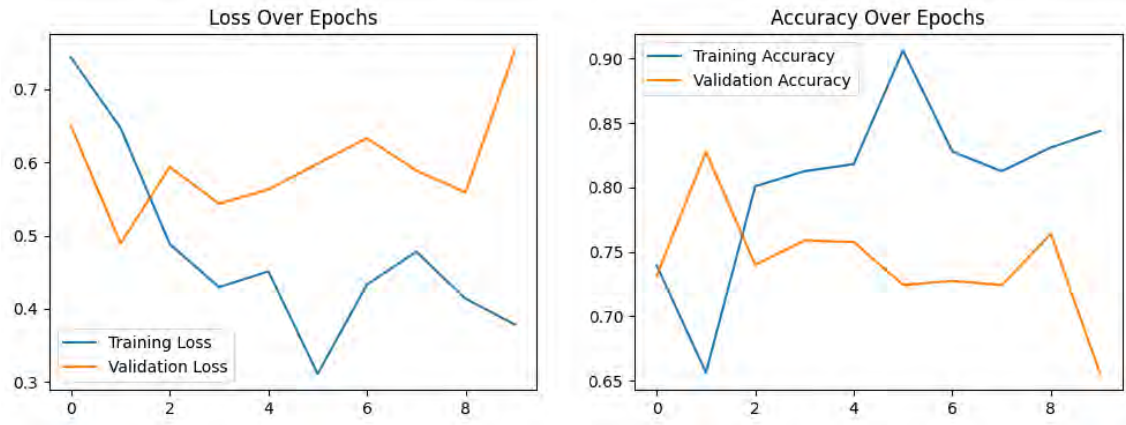


Figure 4.2: InceptionV3 Model Accuracy & Loss Metrics

Class	precision	recall	f1-score	support
HEALTHY	0.99	0.95	0.97	1023
MLN_1_and_MLN_2	0.88	0.91	0.90	796
MSV_2	0.55	0.80	0.65	631
MSV_1	0.70	0.38	0.50	619
Accuracy			0.80	3069
macro avg	0.78	0.76	0.75	3069
weighted avg	0.81	0.80	0.79	3069

Table 4.3: InceptionV3 Model Report.

4.1.3 VGG16

Using the VGG16 Model with 10 epochs, 383 batches and a data split ratio of 2:8, with 80% for training and 20% for validation. An accuracy of 81% was obtained. To facilitate clear understanding, a visual representation has been employed. The following Training and validation accuracy & Loss in figure 4.10 and table 5.2 highlights the key trends and patterns observed in training the dataset with the VGG16.



Figure 4.3: VGG16 Model Accuracy & Loss Metrics

Class	precision	recall	f1-score	support
HEALTHY	0.97	0.96	0.97	1023
MLN_1_and_MLN_2	0.91	0.97	0.93	796
MSV_2	0.57	0.82	0.68	631
MSV_1	0.73	0.36	0.48	619
Accuracy			0.81	3069
macro avg	0.80	0.78	0.77	3069
weighted avg	0.82	0.81	0.80	3069

Table 4.4: VGG16 Model Report.

4.1.4 EfficientNetV2B2

Employing the EfficientNetV2B2 model to the dataset with 10 epochs, 383 batches and a data split ratio of 2:8 was used, with 80% for training and 20% for validation which yielded an accuracy of 91%. To visualize the findings, the following confusion matrix in figure 4.4 and table 4.5 highlight the key trends and patterns observed in training the dataset with the EfficientNetV2B2 model.

Class	precision	recall	f1-score	support
HEALTHY	0.99	0.99	0.99	1023
MLN_1_and_MLN_2	0.95	0.97	0.96	796
MSV_2	0.81	0.85	0.83	631
MSV_1	0.87	0.79	0.82	619
Accuracy			0.92	3069
macro avg	0.90	0.90	0.90	3069
weighted avg	0.92	0.92	0.92	3069

Table 4.5: EfficientNetV2B2 Model Report.

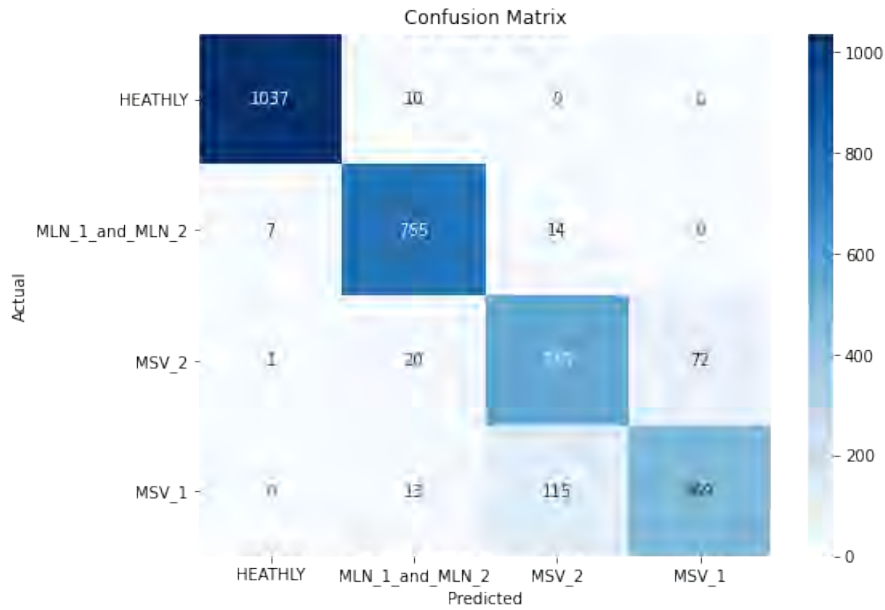


Figure 4.4: EfficientNetV2B2 Model Confusion Matrix

4.1.5 Xception

Training the datasets with Xception Model using 10 epochs, 383 batches and a data split ratio of 2:8, with 80% for training and 20% for validation yielded an accuracy of 89%. To aid understanding, the following training and validation accuracy & loss in figure 4.5, figure 4.6 and table 4.6 highlight the key trends and patterns observed in training the dataset with the Xception model.

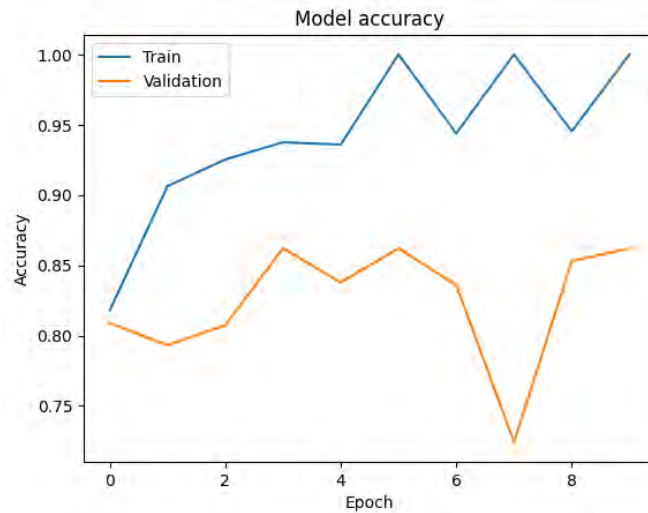


Figure 4.5: Xception Model Accuracy Metric

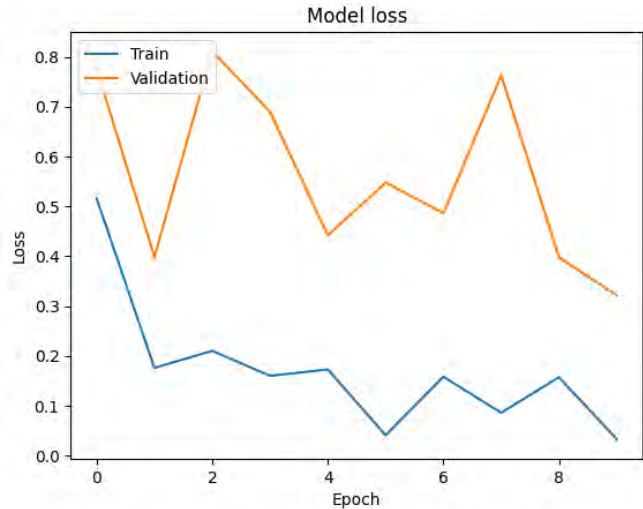


Figure 4.6: Xception Model Loss Metric

Class	precision	recall	f1-score	support
HEALTHY	1.00	1.00	1.00	1023
MLN_1_and_MLN_2	0.99	0.97	0.98	796
MSV_2	0.88	0.55	0.68	631
MSV_1	0.66	0.94	0.78	619
Accuracy			0.89	3069
macro avg	0.89	0.86	0.86	3069
weighted avg	0.91	0.89	0.88	3069

Table 4.6: Xception Model Report.

4.2 Vision Transformer & Hybrid Model’s Performance

Table 4.7 summarizes the hyperparameters for the various Vision Transformer models utilized in the experiment. All models—SWIN, DaViT, MobileViT, MaxViT, and Involution Neural Network (INN) used the Adam optimizer to enhance their training in batches of size 32. The setting for the learning rate was 0.0001 for all models except INN, which had a learning rate of zero. The input size for each model was standardized to 224x224x3, ensuring consistency between models. Every model was trained for thirty epochs using Categorical Crossentropy as the loss function, which is suitable for multi-class classification. These hyperparameter settings were chosen to achieve an equitable balance of learning capability and generalization across models.

Table 4.7: ViT HyperParameter Settings

Parameter	SWIN	DaViT	MobileViT	MaxViT	INN
Batch Size	32	32	32	32	32
Optimizer	Adam	Adam	Adam	Adam	Adam
Learning Rate	0.0001	0.0001	0.0001	0.00001	0
Input Size	224×224×3	224×224×3	224×224×3	224×224×3	224×224×3
Epochs	30	30	30	30	30
Loss Function	Categorical Crossentropy				

4.2.1 SWIN ViT

The SWIN Transformer model was trained and evaluated utilizing an 80:20 split, then optimized with the Adam optimizer at a learning rate of 0.0001 across 30 epochs. The model’s input size have been standardized at 224x224x3, and the batch of size 32. SWIN obtained 93% accuracy in maize disease classification as shown in table 4.8, demonstrating its potential to capture both local and global contextual data using its hierarchical architecture and window-based self-attention mechanism. In figure 4.7 highlights the key trends in patterns observed by SWIN ViT Model.

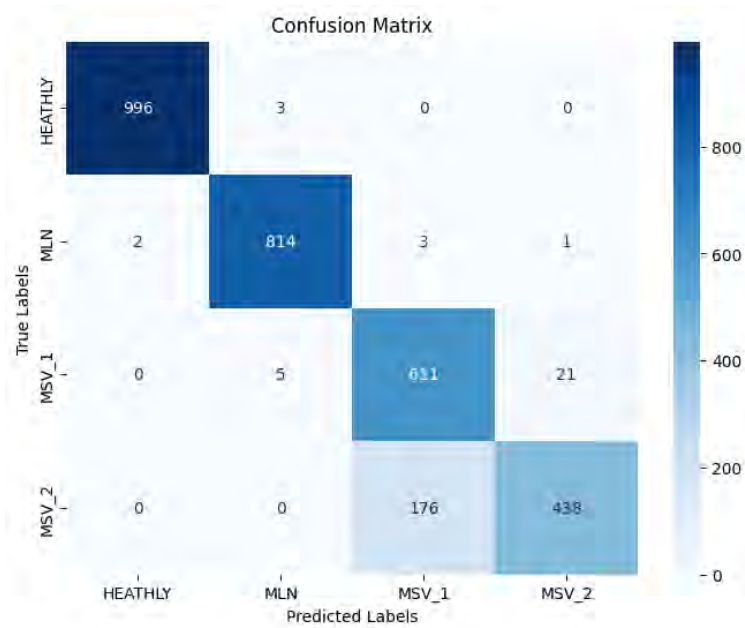


Figure 4.7: Swin ViT Model Confusion Matrix

The Categorical Crossentropy loss function was used for multiclass classification, which is excellent for this task as depicted in figure 4.8 and figure 4.9. With a GPU memory consumption of 1369.42 MB and an average epoch time of 260 seconds, SWIN was computationally intensive but extremely effective, establishing a mix between efficiency and accuracy in dealing with complicated visual patterns.

Class	precision	recall	f1-score	support
HEALTHY	1.00	1.00	1.00	999
MLN	0.99	0.99	0.99	820
MSV_1	0.77	0.96	0.86	637
MSV_2	0.95	0.71	0.82	614
Accuracy			0.93	3070
macro avg	0.93	0.92	0.92	3070
weighted avg	0.94	0.93	0.93	3070

Table 4.8: Swin Transformer Model Report.

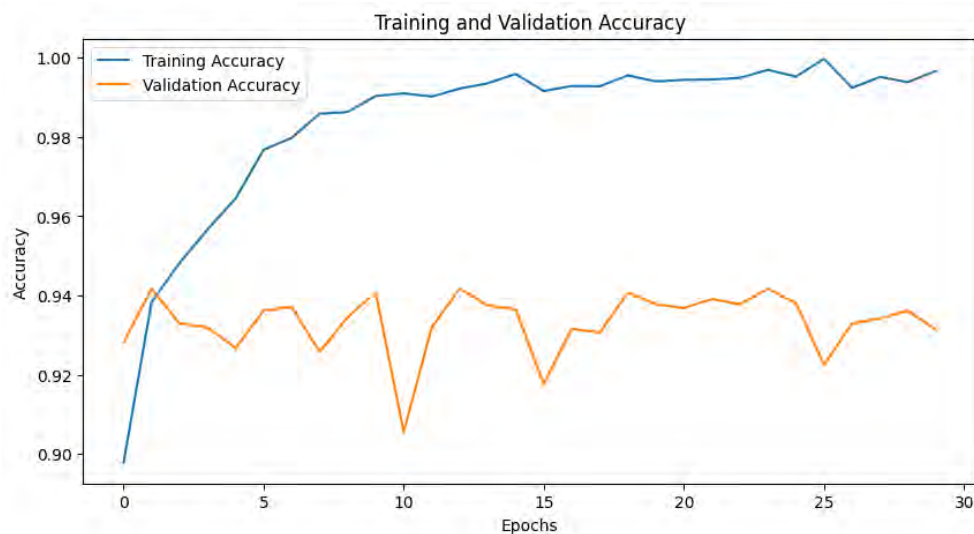


Figure 4.8: SwinViT Accuracy Metric

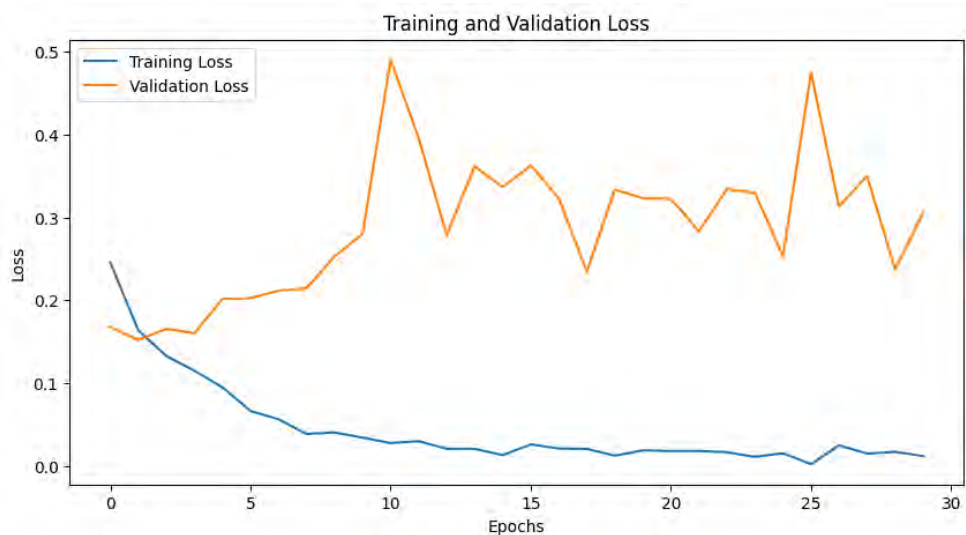


Figure 4.9: SwinViT Loss Metric

4.2.2 DaViT

The DaViT model was employed with an 80/20 split for training and testing, 30 epochs, a batch size of 32, and the Adam optimizer with a learning rate of 0.0001.

The input size was set to 224x224x3, which ensured uniformity between models. DaViT exhibited an impressive 95% accuracy in maize disease classification, matching the best performance shown in table 5.2 and figure 4.10 which depicts the key trends in pattern observed.

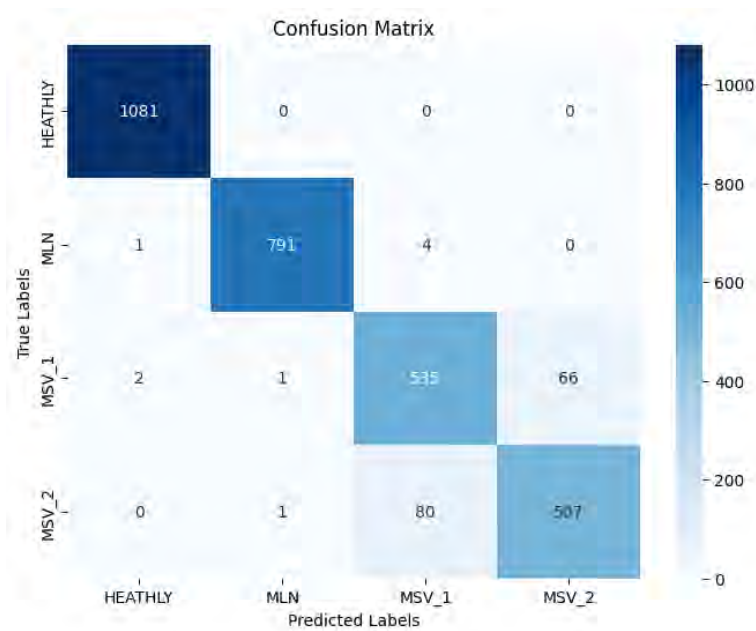


Figure 4.10: DaViT Model Confusion Matrix

Class	precision	recall	f1-score	support
HEALTHY	1.00	1.00	1.00	1081
MLN	1.00	0.99	1.00	796
MSV_1	0.86	0.88	0.87	604
MSV_2	0.88	0.86	0.87	588
Accuracy			0.95	3069
macro avg	0.94	0.94	0.94	3069
weighted avg	0.95	0.95	0.95	3069

Table 4.9: DaViT Model Report.

Its dual-attention technique, which combines global and local feature extraction, was extremely effective in collecting complex patterns, even under difficult situations such as lighting or image orientation changes as shown in figure 4.11. This capability made DaViT ideal for activities that required more in-depth contextual understanding. The model was trained with Categorical Crossentropy as the loss function, which was a good fit for the multi-class classification challenge and contributed to the model's strong performance across the dataset.

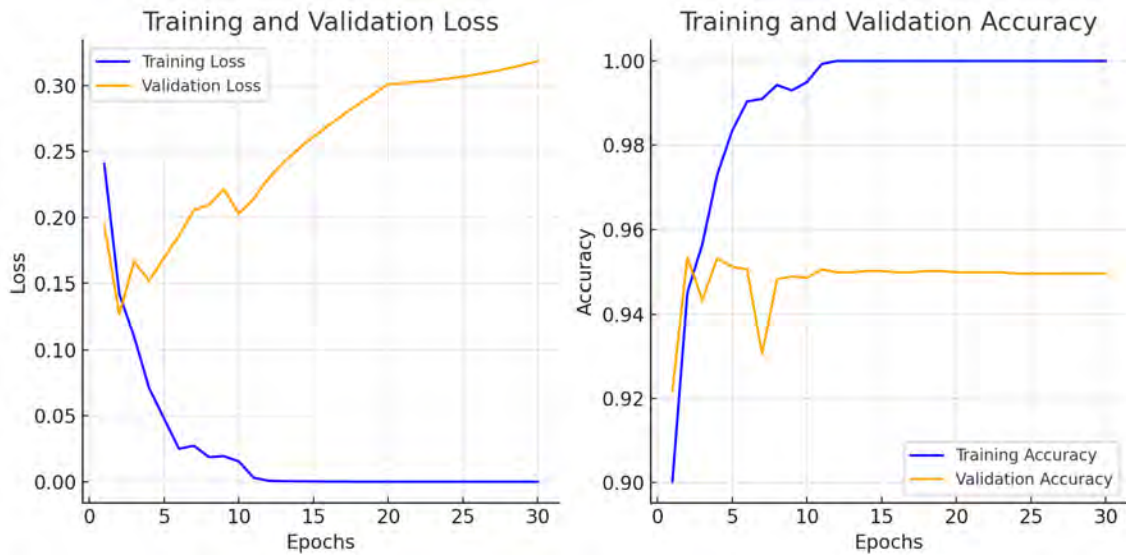


Figure 4.11: DaViT Model Metrics

4.2.3 MobileViT

The MobileViT model was trained with an 80:20 split for training and testing, 30 epochs, and 32 batches. Employed learning rate of 0.0001 and Adam optimizer, the input size have been set to 224x224x3. MobileViT obtained an outstanding 95% accuracy as depicted in table 4.10, placing it among the top-performing models in the research study. The model’s lightweight architecture, which blends convolutional layers with self-attention processes, enabled it to efficiently capture both local and global information, making it very successful at detecting subtle trends in maize disease classification. In figure 4.12 shows the confusion matrix which highlights the patterns and observertion of the model.

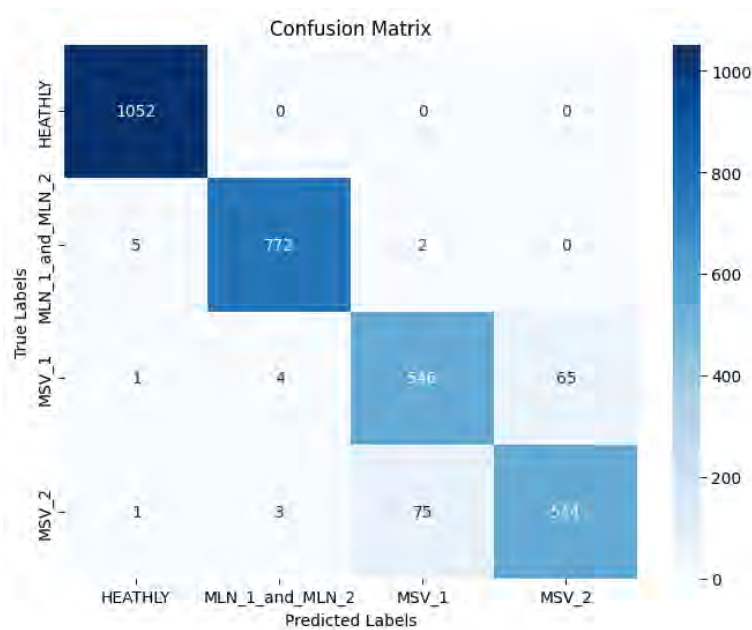


Figure 4.12: MobileViT Model Confusion Matrix

Class	precision	recall	f1-score	support
HEALTHY	0.99	1.00	1.00	1052
MLN	0.99	0.99	0.99	779
MSV_1	0.88	0.89	0.88	616
MSV_2	0.89	0.87	0.88	623
Accuracy			0.95	3070
macro avg	0.94	0.94	0.94	3070
weighted avg	0.95	0.95	0.95	3070

Table 4.10: MobileViT Model Report.

MobileViT’s design also allows it to be deployed on edge devices with low computational overhead—its GPU memory utilization during training was only 128.78 MB, making it perfect for real-time applications. Categorical Crossentropy was used as the loss function to fit the model’s multi-class classification problem.

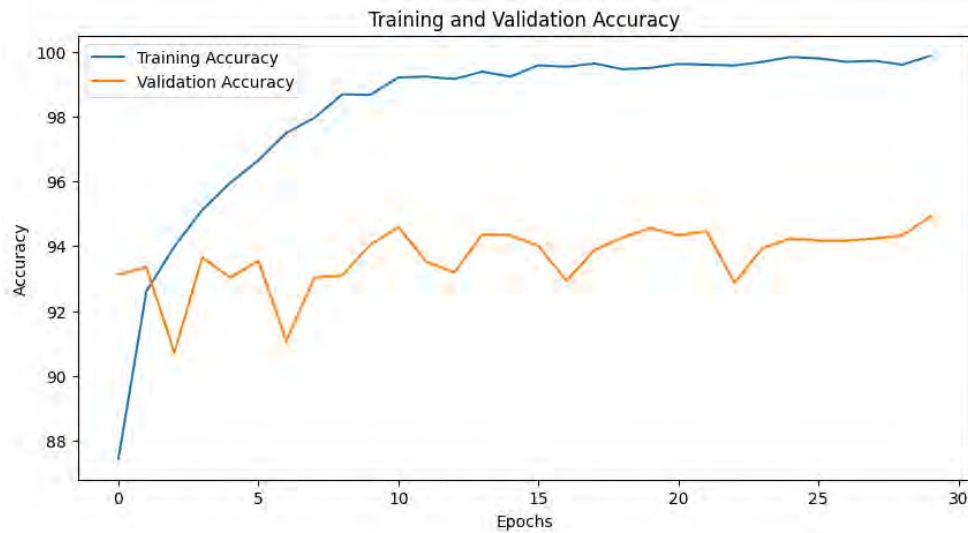


Figure 4.13: MobileViT Accuracy Metric



Figure 4.14: MobileViT Loss Metric

4.2.4 MaxViT

MaxViT model was trained with an 80:20 train-test split, as were the other Vision Transformer models. The training lasted 30 epochs, with a batch of size 32, utilizing the Adam optimizer and a learning rate of 0.0001. To provide consistent results across all models, input images were scaled to 224x224x3. MaxViT, which combines local and global attention mechanisms, was tuned with the Categorical Crossentropy loss function, the model used a GPU Memory of 1865.52 MB during training for approximately 600 seconds per epoch. Suitable for various class classification tasks. for understanding the figure 4.18 highlights the key trends and patterns observed.

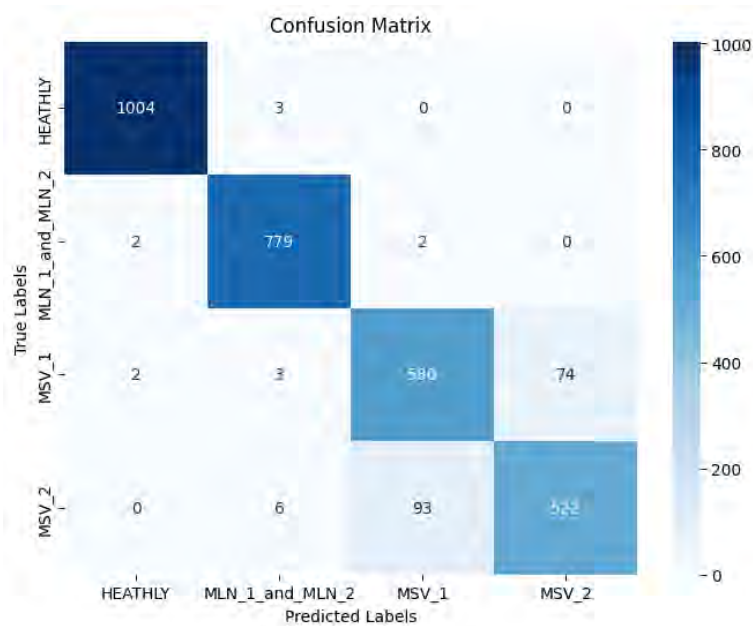


Figure 4.15: MaxViT Model Confusion Matrix

Class	precision	recall	f1-score	support
HEALTHY	1.00	1.00	1.00	1007
MLN	0.98	0.99	0.99	789
MSV_1	0.86	0.88	0.87	659
MSV_2	0.88	0.84	0.86	621
Accuracy			0.94	3070
macro avg	0.93	0.93	0.93	3070
weighted avg	0.94	0.94	0.94	3070

Table 4.11: MaxViT Model Report.

The model’s design uses a unique grid and block attention to extract rich feature representations, allowing it to perform well across a wide range of datasets as you can see it in table 4.11. This balanced hyperparameter and model design combination improved classification performance while remaining efficient in feature extraction and generalization.

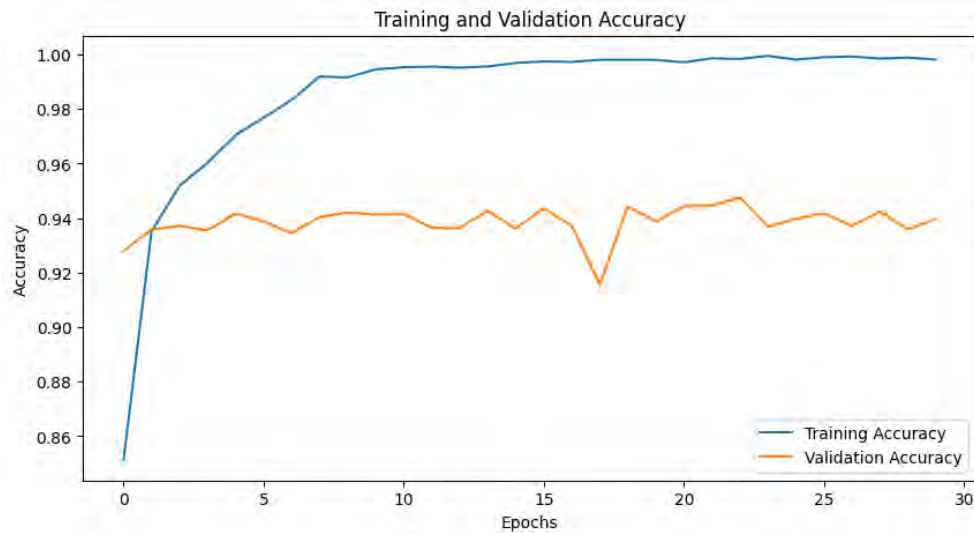


Figure 4.16: MaxViT Accuracy Metric



Figure 4.17: MaxViT Loss Metric

4.2.5 INN

In my study, we employed the Involutional Neural Network (INN) with an 80:20 for training and testing split to ensure efficient data distribution for both training and evaluation. The model was trained over 30 epochs with a batch size of 32, utilizing the Adam optimizer to efficiently update weights. Unlike other models, the INN’s learning rate was set to 0, allowing it to alter dynamically during training as shown in figure 4.19. Input images were standardized to 224x224x3, ensuring consistency across all models as shown in figure 4.5. The loss function used was Categorical Crossentropy, which was perfect for the various-class classification task.

Class	precision	recall	f1-score	support
HEALTHY	0.99	1.00	0.99	5117
MLN	0.99	0.97	0.98	3980
MSV_1	0.85	0.78	0.81	3157
MSV_2	0.80	0.87	0.83	3095
Accuracy			0.92	15349
macro avg	0.91	0.90	0.90	15349
weighted avg	0.92	0.92	0.92	15349

Table 4.12: INN Model Report

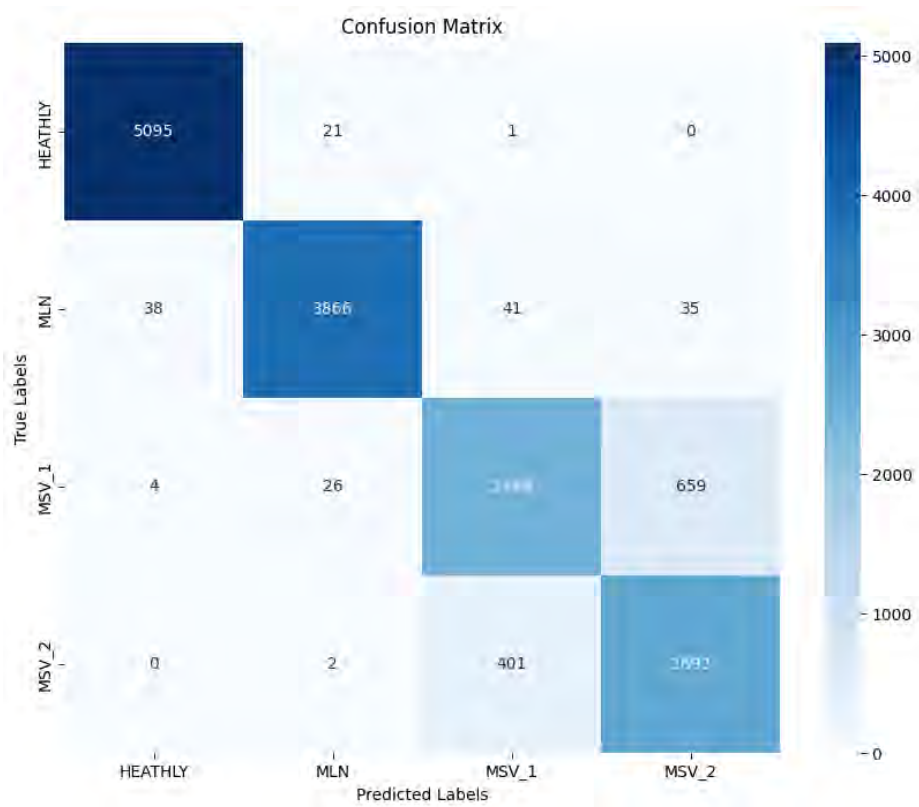


Figure 4.18: INN Model Confusion Matrix

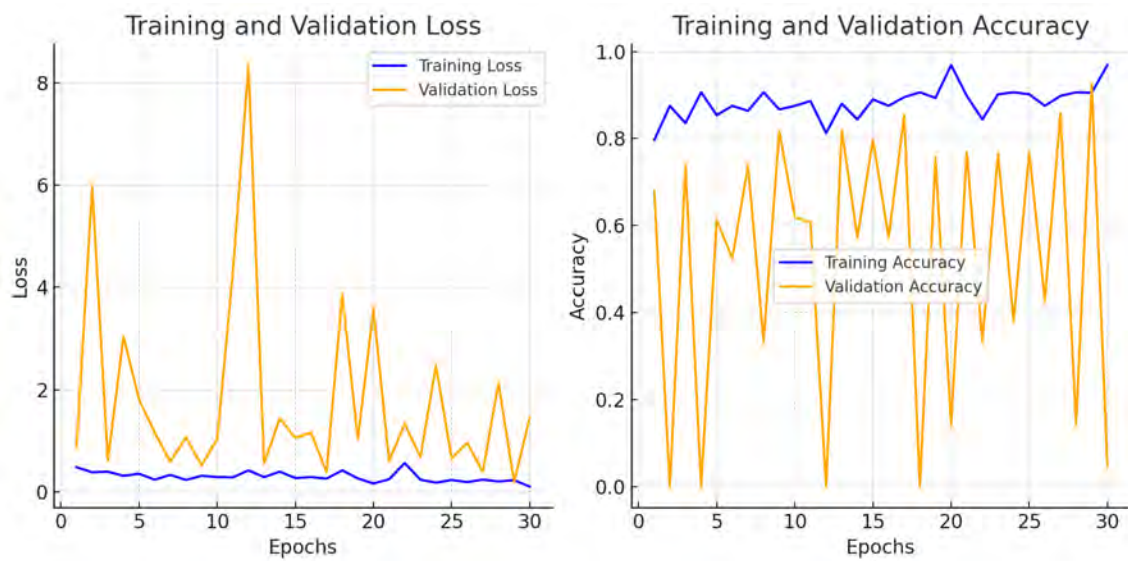


Figure 4.19: INN Model Metric

This configuration intended to balance learning capacity which yielded 92% as depicted in table 4.6 and generalization while exploiting INN’s novel architectural design, which improves feature extraction by focusing on local features via involution operations.

4.2.6 Fusion Model

To improve maize disease classification performance, a fusion model was built utilizing two pre-trained Vision Transformer models: MobileViT and DaViT. Both models used 30 epochs for training, utilizing the Adam optimizer which had a learning rate of 0.0001 and weight decay of 1e-5. For multi-class classification, the fusion model's criteria was set to nn.CrossEntropyLoss, and training was done using a batch of size 64. The fusion architecture brings together the complementing qualities of both models: MobileViT, which is noted for its lightweight design and computational efficiency, and DaViT, which has powerful attention mechanisms that excel at processing complicated patterns. The final layers of both models are combined in the fusion to form a strong feature representation that is then passed to the classifier which allows our model to get an accuracy of 96 % as shown in table 4.13. Moreover, our model was able to classify well images in their distinct classes as depicted in fig 4.20.

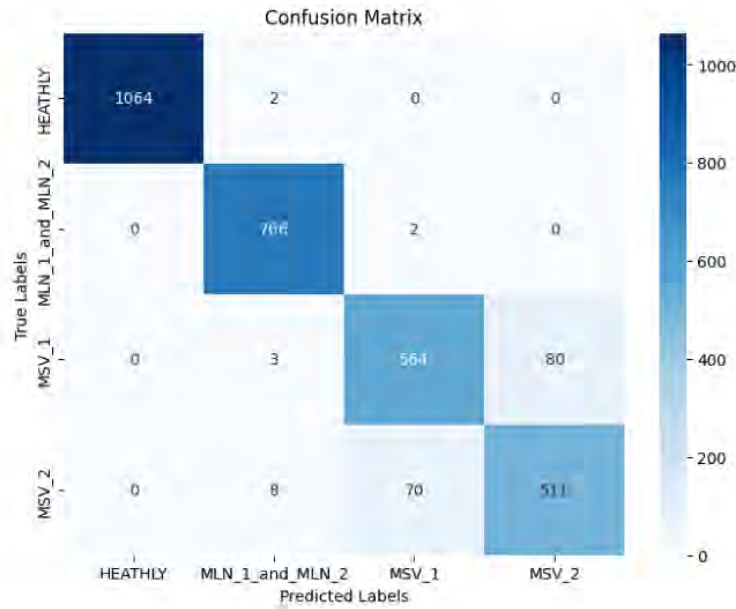


Figure 4.20: Fusion Model Confusion Matrix

Class	precision	recall	f1-score	support
HEALTHY	1.00	1.00	1.00	1066
MLN	0.98	1.00	0.99	768
MSV_1	0.90	0.89	0.89	647
MSV_2	0.89	0.89	0.90	589
Accuracy			0.96	3070
macro avg	0.94	0.95	0.95	3070
weighted avg	0.95	0.96	0.96	3070

Table 4.13: Fusion Model Report

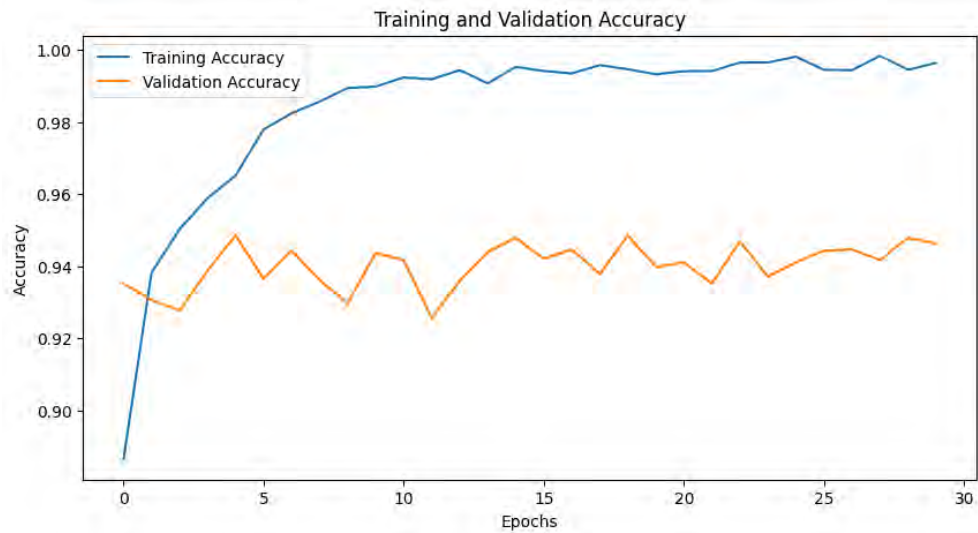


Figure 4.21: Accuracy Metrics for Fusion Model

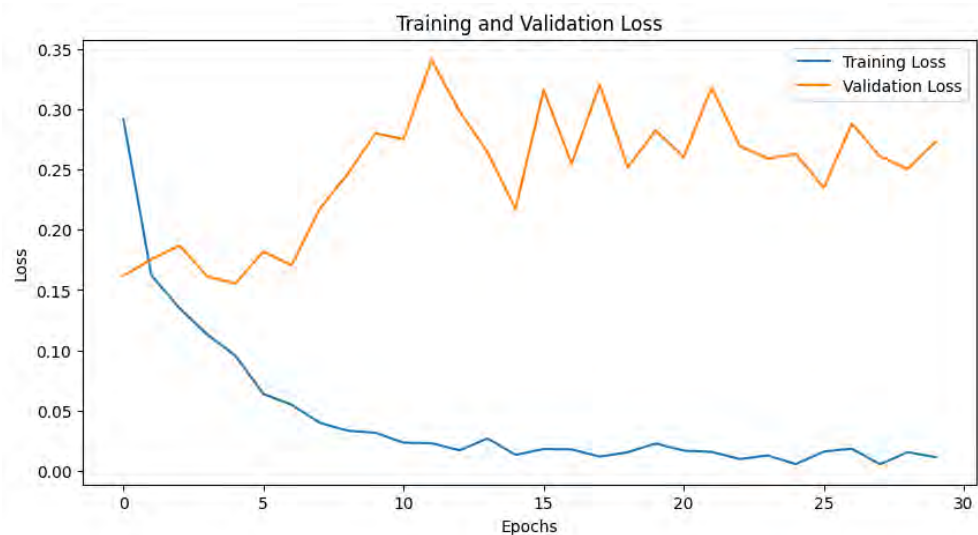


Figure 4.22: Loss Metrics for Fusion Model

This fusion model demonstrates stable GPU memory usage during the training process using only 865.78 MB, demonstrating efficient resource utilization with no significant spikes or fluctuations across the 30 epochs, and achieving greater accuracy in training and validation than loss, as depicted in fig 4.21 and fig 4.22. This hybrid approach increased model robustness and classification accuracy.

4.3 Comparative Analysis

In this study, Transfer Learning (TL) and Vision Transformer (ViT) models were used to classify maize leaf diseases. The best model in the TL category obtained 92% accuracy, while the bottom-performing model achieved 80%. In contrast, Vision Transformer models performed better, with the greatest accuracy of 95% and even the lowest-performing ViT model reaching 92% as shown in figure 4.23. This demonstrates that Vision Transformers outperform Transfer Learning models in most situations, particularly when dealing with complicated visual patterns and extracting global characteristics.

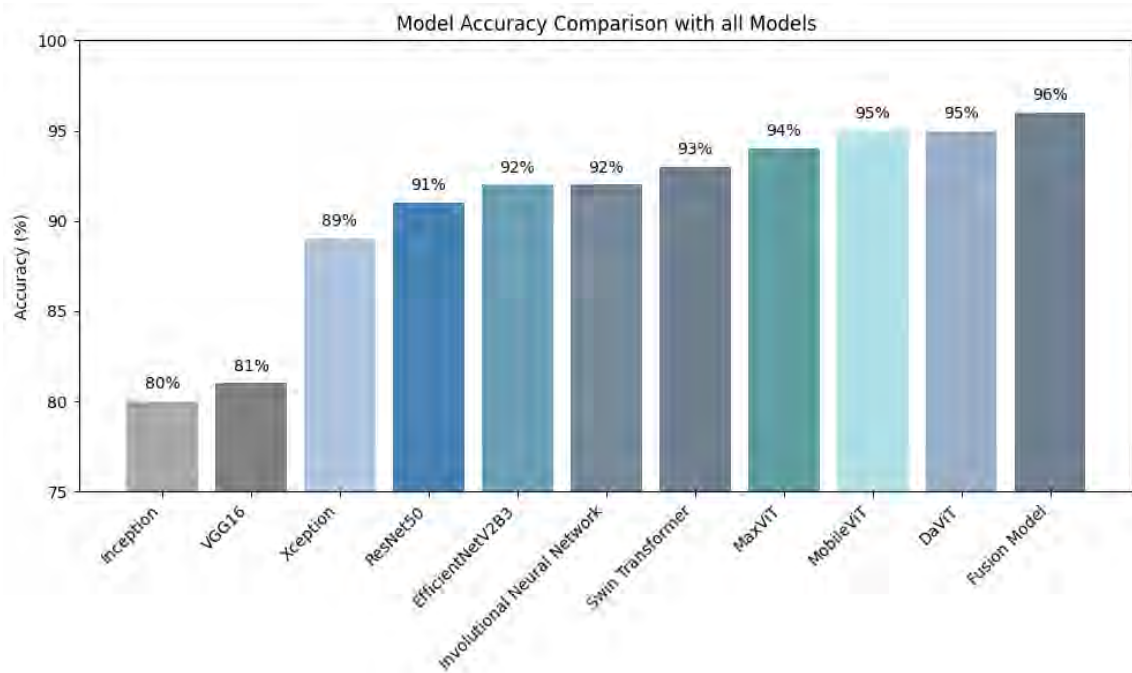


Figure 4.23: All Models Comparison

Our fusion model outperformed single models with a prediction accuracy of 96%. This highlights ViT models' accuracy and ability to handle complicated visual data. The ensemble prediction is based on a robust and complementary fusion of the models DaViT and MobileViT.

Table 4.14: Transfer Learning vs. Vision Transformers

Aspect	Transfer Learning Models	Vision Transformer Models
Best Accuracy	92% (EfficientNetV2B2)	95%(MobileViT, DaViT)
Lowest Accuracy	80% (InceptionV3)	92% (INN)
Parameters (Millions)	25-30M	40-50M
Architecture	CNN-based, feature extraction through pre-trained models	Self-attention mechanisms, better at capturing global dependencies
Performance	Good, but less effective for complex patterns	Superior in feature extraction and classification tasks
Complexity	Lower computational cost, easier to deploy	Higher complexity but better for large datasets
Global Feature Handling	Limited	Excellent

From table 4.14 highlights a detailed comparison, demonstrating that Vision Transformers (ViTs) outperform Transfer Learning models in accuracy and global feature handling, although using more processing resources. The parameter count is likewise higher for ViTs, indicating their complexity in feature extraction.

The study we conducted intended to accurately classify maize diseases using sophisticated deep learning models, with an emphasis on Vision Transformer (ViT) architectures. Models were selected using specified criteria to ensure best performance in terms of accuracy, computational efficiency, scalability, and real-world relevance to agricultural diagnostics. The following are the major criteria that guided the selection of ViT models for maize disease classification.

Table 4.15: Selection Criterias for Vision Transformer(ViT) Models

Criteria	Swin ViT	MaxViT	MobileViT	DaViT	INN
Architectural Focus	Shifted window attention for local-global feature capture	Multi-axis attention for local-global feature fusion	Hybrid CNN-Transformer architecture for lightweight performance	Dual attention mechanism for comprehensive global attention	Local feature learning using in-volution operation

Scalability	Highly scalable for high-resolution images	Efficient in handling varying resolutions	Optimized for low-resource environments (mobile/edge devices)	Scales well across resolutions, suited for complex datasets	Primarily focuses on local patterns, may require modifications for larger-scale data
Computational Efficiency	Moderate, suitable for high-end GPUs	High efficiency despite complex architecture	Lightweight, designed for low-memory devices	Moderate, suitable for large datasets	Efficient in capturing local features but may be slower on global features
Accuracy for Fine-Grained Classification	Excellent for distinguishing subtle variations	Strong performance due to multi-axis attention	Good performance, slightly lower than heavier models	Strong attention mechanism enhances classification accuracy	Performs well in capturing local differences but less effective globally
Multi-Scale Feature Learning	Yes, via hierarchical representation	Yes, with efficient multi-axis mechanism	Limited, focuses more on lightweight performance	Yes, with dual attention mechanism	Primarily focused on local features rather than multi-scale integration
Real-Time Deployment	Requires higher computational resources	Suitable for high-end deployment	Best suited for real-time edge deployment	Requires higher resources for deployment	Can be deployed in real-time with optimizations
Explainability and Interpretability	Moderate, due to window attention mechanism	High, captures complex interrelations	Limited, optimized for efficiency over interpretability	High, due to attention mechanisms	High explainability, especially in local feature interpretation

We chose ViT models to reconcile high diagnostic accuracy with computational efficiency and scalability as depicted in table 4.15. SWINViT, MaxViT, and DaViT offer robust systems for capturing both global and local features, which are critical for distinguishing between distinct maize diseases. MobileViT provides a lightweight option for real-time deployment in limited resource situations, but the Involucional Neural Network adds interpretability, making it appropriate for explainable AI applications in agriculture. Together, these models form a comprehensive toolkit for achieving high accuracy and efficiency in maize disease diagnosis.

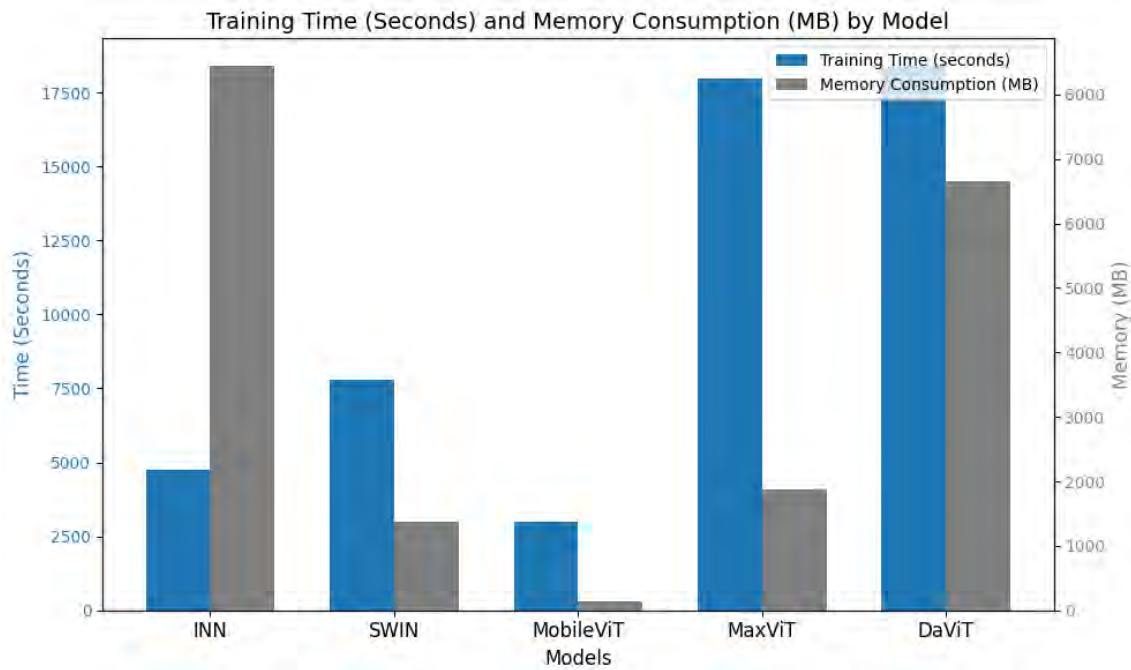


Figure 4.24: Training time Vs Memory usage

In summary, the comparative analysis of Vision Transformer (ViT) models demonstrates the various capabilities and trade-offs in performance and resource efficiency. Each model was chosen based on its architectural strengths, computing requirements, and application to the maize disease classification task. The study emphasizes the distinctions between standard Transfer Learning methods and more advanced ViT models, notably in terms of capturing complex patterns and fine-grained features. While some models excel in accuracy and attention processes, others are more scalable and deployable. Furthermore, the comparison of training times and memory usage as depicted in figure 4.24 explains the practical aspects of real-time applications. This review sheds light on these models' potential for improving maize disease detection and classification, hence helping to the development of more efficient and accurate diagnostic tools in the agricultural realm.

Chapter 5

Hardware Deployment

As the demand for real-time image classification is increasing, especially in the agriculture sector, there is a significant challenge in deploying a machine learning model which is computationally expensive on edge devices. Deep learning models for instance require hardware with high performance to process data efficiently. However, in real-world scenarios where power, space, and computational resources are limited, deploying such a model remains an obstacle. Raspberry Pi is one of the portable and cost-effective edge devices widely used for real-time processing. However, It has limited CPU power and memory, which restricts the deployment of large and complex models typically trained for image classification. By optimizing and compressing the hybrid model, the problem of deploying highly accurate, pre-trained deep-learning models on the Raspberry Pi for high image classification will be addressed. Table 5.1 shows the hardware specs of the Raspberry Pi used.

Specification	Details
Processor	Broadcom BCM2711, Quad-core Cortex-A72 @ 1.8GHz
RAM	1GB, 2GB, 4GB, or 8GB LPDDR4-3200 SDRAM
Wireless	2.4 GHz & 5.0 GHz IEEE 802.11ac, Bluetooth 5.0
Ethernet	Gigabit Ethernet
USB Ports	2 × USB 3.0; 2 × USB 2.0
GPIO Header	40-pin GPIO header (backwards compatible)
HDMI Ports	2 × micro-HDMI® (up to 4kp60)
Display Port	2-lane MIPI DSI
Camera Port	2-lane MIPI CSI
Audio/Video Port	4-pole stereo audio & composite video
Video Decoding	H.265 (4kp60), H.264 (1080p60 decode, 1080p30 encode)
Graphics API	OpenGL ES 3.1, Vulkan 1.0
Storage	Micro-SD card slot
Power Supply	5V DC via USB-C (minimum 3A)
Power Supply via GPIO	5V DC via GPIO header (minimum 3A)
Power over Ethernet	PoE enabled (requires separate PoE HAT)
Operating Temperature	0 – 50 degrees C ambient
Power Supply Note	2.5A supply can be used if USB peripherals consume < 500mA

Table 5.1: Raspberry Pi Specifications

5.1 Setup

Hardware Requirements	
Component	Description
MicroSD card	32GB
Power supply	For Raspberry Pi 4
Micro-HDMI cable	Connect the Raspberry Pi to the screen
Mobile or desktop device	To interact with the Flask web app on the Raspberry Pi
Software Requirements	
Component	Description
Operating System	Raspberry Pi OS
Python 3	Installed on Raspberry Pi
Flask	Python microframework for web applications
Torch & TorchVision	For model inference and image processing
PIL (Pillow)	Image processing library for Python
Timm	PyTorch Image Models (for loading pre-trained models)

Table 5.2: Hardware and Software Requirements

Every hardware in Table 5.2 is connected to the Raspberry Pi. Next, the Raspberry Pi's microSD port is inserted with a microSD card that contains the OS installed on it. Using the terminal, `sudo apt update` and `sudo apt upgrade` are then typed to update and upgrade the system respectively. Python 3 is then installed using `sudo apt install python3`, the virtual environment is created and updated using `python -m my_venv/bin/activate` and `source my_venv/bin/activate`. `pip` is used to install all the necessary libraries(Flask, PyTorch, TorchVision, Pillow, Timm).



Figure 5.1: Raspberry Pi Setup

5.2 Model Compression and Deployment

To deploy the ensemble model on the Raspberry Pi 4 efficiently, only the essential parameters has to be saved using `torch.save` function which reduces the model size. The pre-trained model was transformed into a serialized format, and the weights were optimized. The reduced model is capable of running on the CPU of the Raspberry Pi without overwhelming its processing capabilities.

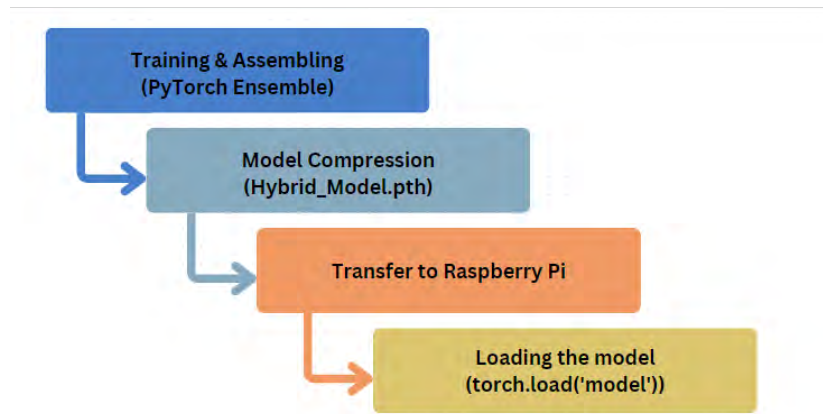


Figure 5.2: Model Deployment

5.3 System Architecture

At the core of the system is Flask, a web application that offers users a simple interface for uploading a maize image from their device. The Raspberry Pi processes and classifies the image. The results are then sent to the user's device. The flask application includes the `app.py`, and it handles loading the model, image uploads, and the classification logic. The HTML files for the interface are contained in the `templates` folder of the Flask application, and the uploaded images are kept in the `upload` folder. The image classification pipeline follows multiple stages: the image uploaded is resized to 224 by 224 pixels and normalized to meet the model input requirements, and then for inference, it passes through the ensemble model, and then finally, the highest probability class is returned to the user. With the Flask app directory and all the necessary files inside it, once the server is running, users can access the interface through the IP address of the Raspberry Pi.

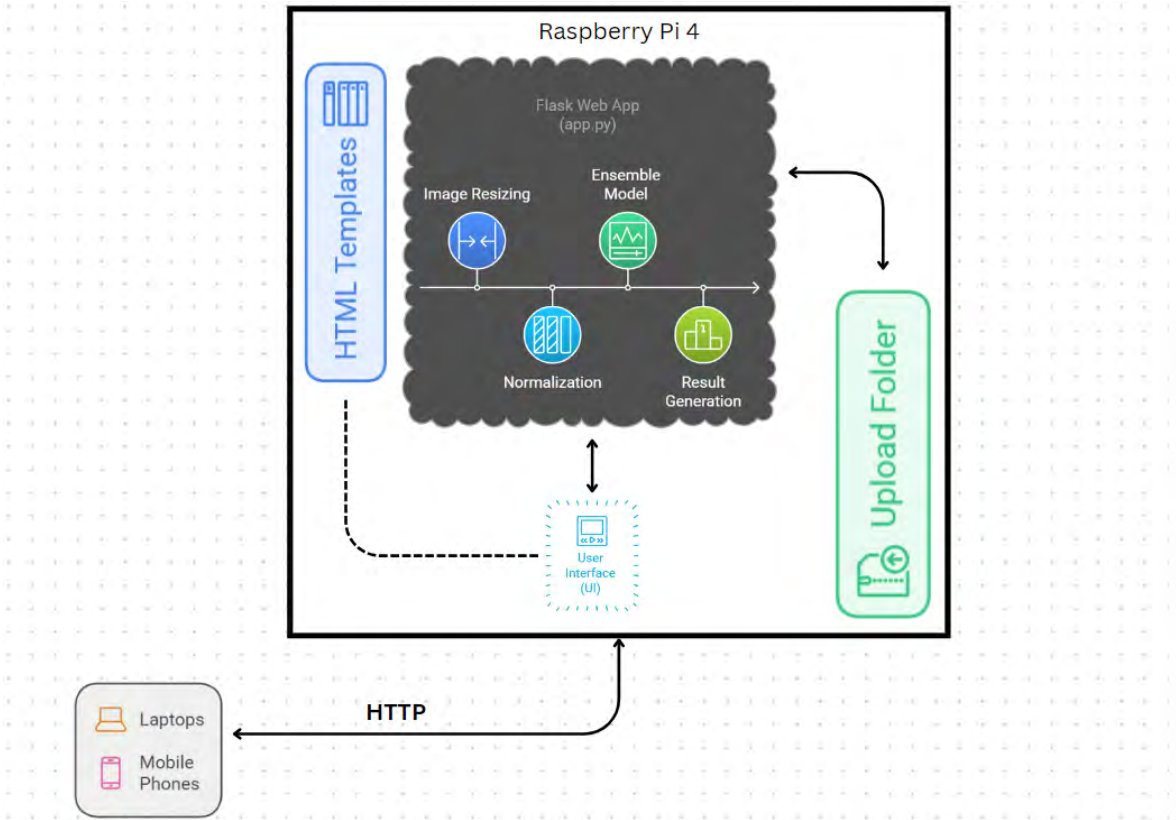


Figure 5.3: Hardware System Architecture

Algorithm 1 Image Upload and Classification System

- 1: $user_image \leftarrow \text{ReceiveUploadedImage}()$
 - 2: $saved_image_path \leftarrow \text{SaveImage}(user_image, upload_folder)$
 - 3: $preprocessed_image \leftarrow \text{PreprocessImage}(saved_image_path, \text{Size: } (224, 224))$
 - 4: $model \leftarrow \text{LoadModel}(model_path)$
 - 5: $output \leftarrow \text{ModelForwardPass}(model, preprocessed_image)$
 - 6: $predicted_class_index \leftarrow \text{GetPredictedClass}(output)$
 - 7: $predicted_class_name \leftarrow \text{MapClassIndexToName}(predicted_class_index, class_names)$
 - 8: $\text{DisplayPredictedClass}(predicted_class_name)$
-

5.4 System Demonstration

In this section, the system's operation is demonstrated in Figure 5.4 which starts with the user uploading an image through the Flask web and concludes with the user seeing the results.

5.5 Inference Results

The performance of the system was evaluated using some 10 images each for Healthy, MSV, and MLN. The metrics evaluated include the confidence score, inference time, CPU usage and memory usage. Figure 5.5 contains the inference performance met-

rics and Table 5.3 summarizes the average performance metrics across 10 images each for Healthy, MSV, and MLN.

Inference Results for Each Class				
Class	Confidence Score (%)	Inference Time (s)	CPU Usage (%)	Memory Usage (MB)
Healthy	99.99	2.13	54.10	0.96
MSV	99.98	2.92	36.80	2.20
MLN	100.00	3.38	40.50	1.89

Table 5.3: Inference Performance Metrics for Each Class

The confidence score of the deployed ensemble model was 99.95% for healthy images. This is an indication of high reliability in distinguishing healthy samples from diseased ones. The model’s inference time was 2.15 seconds for the Healthy class, 2.85 seconds for MSV, and 3.42 seconds for MLN. For CPU usage, Healthy maize images had the highest CPU usage followed by MSV and MLN. The memory usage was high for MSV possibly due to the complexity of MSV images. Despite all the slight variations in the inference time, CPU, and memory usage, The performance of the system remains efficient with CPU resources.

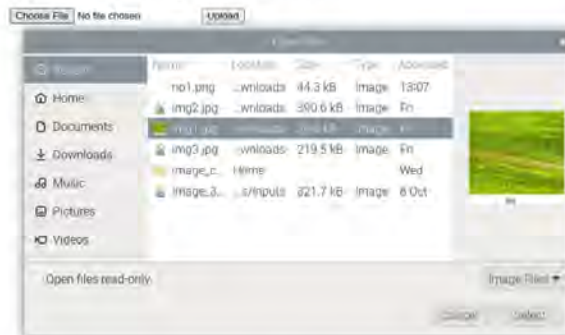
In summary, this chapter gives the steps necessary to deploy the ensemble model for maize disease detection using a Raspberry Pi, focusing on challenges in running resource-intensive models in low-power environments. This chapter also outlines the hardware setup, showing the important elements of Raspberry Pi 4 and the software required. With respect to optimization, the model is compressed using torch.save to reduce computational load. The design of the system shall be web-based, built on Flask, which classifies images in real-time and allows users to upload images to diagnose diseases. Finally, the performance of the system is demonstrated with very efficient use of CPUs and memory, performing accurate and timely inferences across classes of diseases.

Upload an Image for Classification

Choose File No file chosen Upload



Upload an Image for Classification



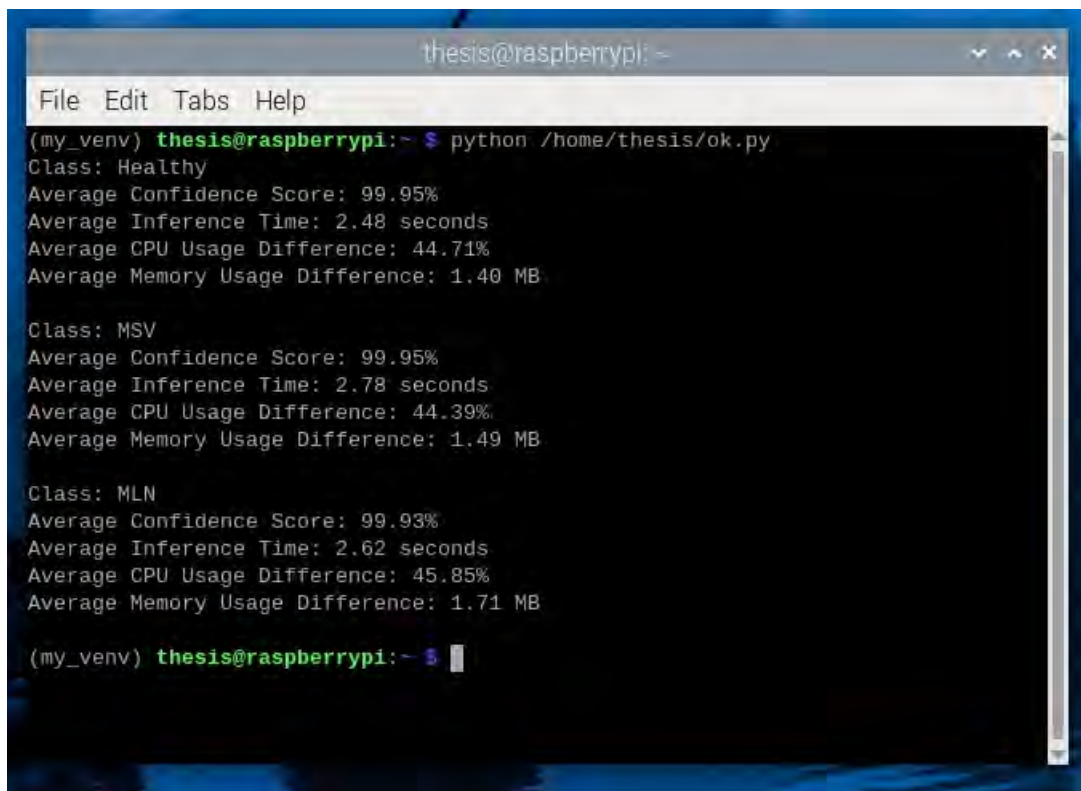
Classification Result

Predicted Class: MLN

Confidence Score: 99.9996542930603%

[Back to Upload](#)

Figure 5.4: System Demonstration



```
thesis@raspberrypi: ~  
File Edit Tabs Help  
(my_venv) thesis@raspberrypi: ~ $ python /home/thesis/ok.py  
Class: Healthy  
Average Confidence Score: 99.95%  
Average Inference Time: 2.48 seconds  
Average CPU Usage Difference: 44.71%  
Average Memory Usage Difference: 1.40 MB  
  
Class: MSV  
Average Confidence Score: 99.95%  
Average Inference Time: 2.78 seconds  
Average CPU Usage Difference: 44.39%  
Average Memory Usage Difference: 1.49 MB  
  
Class: MLN  
Average Confidence Score: 99.93%  
Average Inference Time: 2.62 seconds  
Average CPU Usage Difference: 45.85%  
Average Memory Usage Difference: 1.71 MB  
  
(my_venv) thesis@raspberrypi: ~ $ █
```

Figure 5.5: Inference performance Metrics

Chapter 6

Explainable Artificial Intelligence (XAI)

XAI is a technique used by AI experts to test deep learning (DL) models [33]. It provides the necessary clarity to understand the complex operations of the model and explains the reasons and methods behind the model's output [30]. In this paper, we have used the XAI technique for maize disease detection using a proposed fusion model that includes saliency map, Grad-CAM, LIME, and SHAP. This model combines MobileVit with the DaVit model, providing increased accuracy for object classification.

6.1 Saliency Map

The saliency map, one of the most prominent methods of XAI, shows which portion of the input image is most important for the model's predictions. This is especially relevant in image classification tasks, where it provides a more intuitive explanation of the model's behavior by visually depicting which regions of the input image are the most weighted in the prediction. In order to generate a saliency map, the technique calculates the gradient of the final output with respect to the input image. The numerical output indicates how much the perturbation in regions of the image changes the prediction of the model, and the bigger the area, the more relevant it is. Therefore, this technique acts as a mediator to explain the internal workings of the model from the perspective of the human being, whether the model is predicting based on irrelevant features or necessary aspects.

The saliency map for the maize crop disease detection has been shown in Figures 6.1, 6.2, 6.3 and, 6.4. Remember these figures reproduce in three parts, the first sub-image (left) is the original input maize leaf image. The second sub-image (center) shows the salience map, where red areas describe the most relied on areas by the model in making the prediction. Such red areas are intense to proportions of the classifications that these regions were given towards the end. Finally, the third sub-image (right) shows the saliency map with heatmap overlay. The circles overlay show that the model appears to be centering on specific visual features, including colors and textures, having some abnormality characteristic of the disease symptoms.

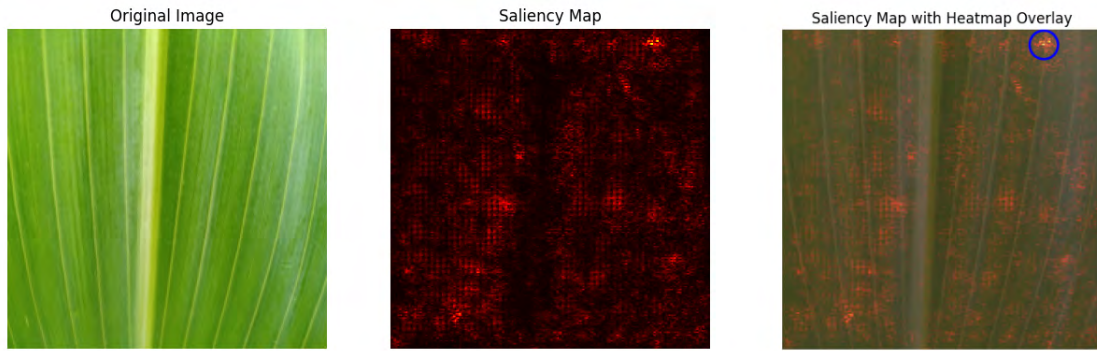


Figure 6.1: Saliency map results for a healthy maize crop using fusion model: The original image (left), saliency map (center), and saliency map with heatmap overlay (right).

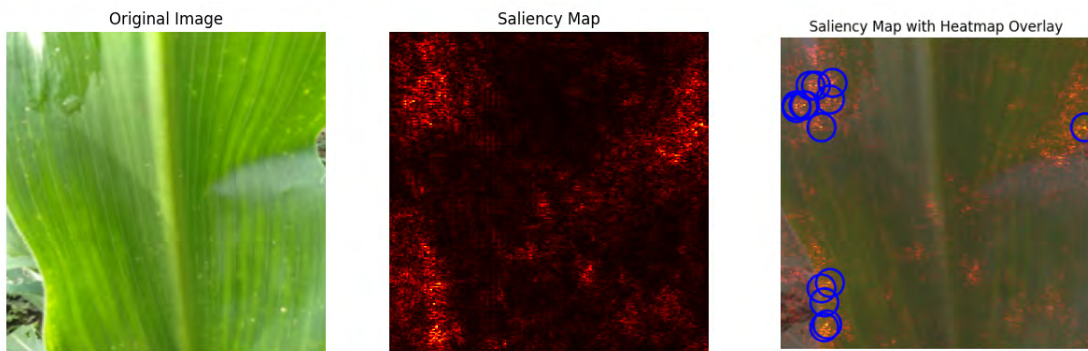


Figure 6.2: Saliency map results for a MLN maize crop using fusion model: The original image (left), saliency map (center), and saliency map with heatmap overlay (right).

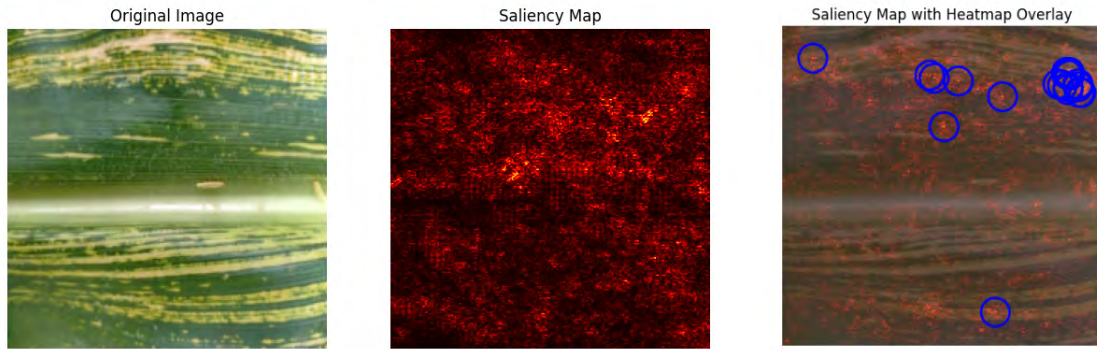


Figure 6.3: Saliency map results for a MSV_1 maize crop using fusion model: The original image (left), saliency map (center), and saliency map with heatmap overlay (right).

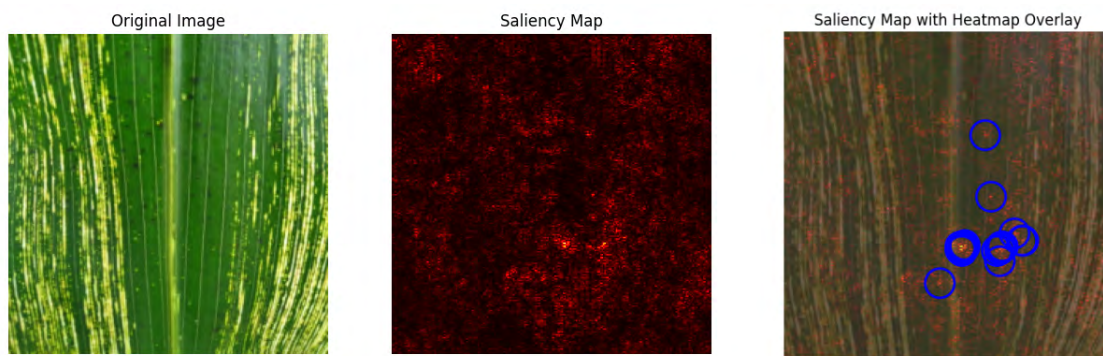


Figure 6.4: Saliency map results for an MSV_2 maize crop using the fusion model: the original image (left), the saliency map (center), and the saliency map with heatmap overlay (right).

6.2 Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) is a visualization technique that helps explain and understand the predictions of the model [46]. It produces class discriminant localization maps, showing which parts of the input image are most valuable in the model’s classification decisions.

For this study, Grad-CAM is essential to enhance the interpretation of the predictions of the proposed fusion model. This allows us to see which regions of maize leaf images the model identifies as indicative of disease, thus validating its focus on relevant features. Grad-CAM helps identify potential misclassifications and facilitates collaboration between data scientists and agronomists. Figure below 6.5, 6.6, 6.7 and, 6.8 below shows the result of the Grad-CAM XAI method with the fusion model.

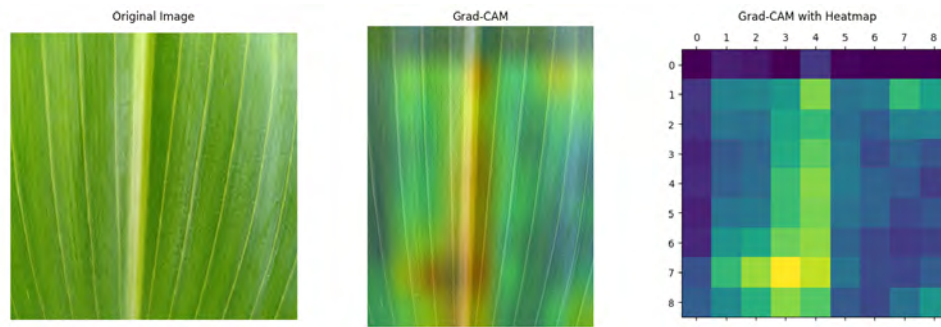


Figure 6.5: Grad-CAM results for a healthy maize crop using fusion model: the original image (left), Grad-CAM (center), and Grad-CAM with heatmap (right).



Figure 6.6: Grad-CAM results for a MLN maize crop using fusion model: the original image (left), GradCAM (center), and Grad-CAM with heatmap (right).

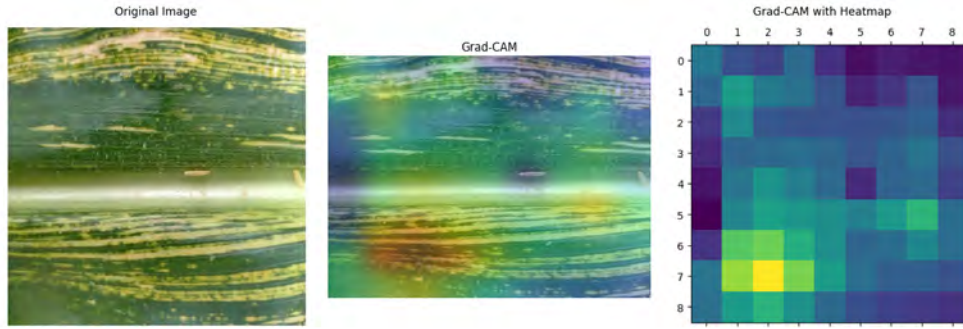


Figure 6.7: Grad-CAM results for a MSV_1 maize crop using fusion model: the original image (left), Grad-CAM (center), and Grad-CAM with heatmap (right).

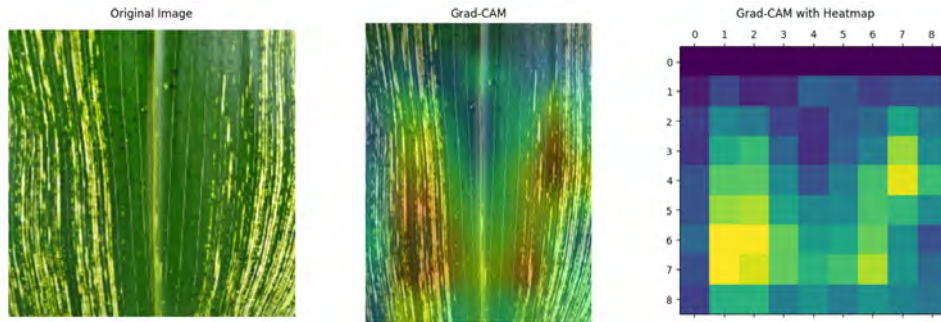


Figure 6.8: Grad-CAM results for a MSV_2 maize crop using fusion model: the original image (left), Grad-CAM (center), and Grad-CAM with heatmap (right).

6.3 LIME

Local Interpretable Model-Agnostic Explanations (LIME) is a method used to interpret the predictions of complex models by approximating them with locally simpler, interpretable models. In image classification tasks, LIME divides an image into superpixels and perturbs them to see how each part affects the model's prediction. It then highlights the most influential areas. In our maize disease detection study, LIME is highly important for detecting which parts of maize leaf images the model is focusing on, ensuring that the model's decisions are based on relevant features.

Figures below 6.9, 6.10, 6.11, and, 6.12 shows a visualization of the explainability results using LIME for a maize disease detection model, with an original maize leaf image, a LIME mask, and a LIME heatmap. The first image (left) in each Figure shows the raw maize leaf input used by the fusion model for classification. Similarly, The second image (center) shows the regions of the original image that LIME identifies as important for the model's decision-making. The yellow regions highlight the segments that contribute most significantly to the classification decision. The scattered nature of the highlighted areas indicates the model's sensitivity to specific leaf texture and patterns, which correlate with disease symptoms. However, the third image (right) incorporates a heatmap to visually represent the importance of different segments. In this context, LIME provides an intuitive way to understand the deep learning model's focus by isolating key areas that contribute most to the prediction, ensuring interpretability and facilitating validation by domain experts.

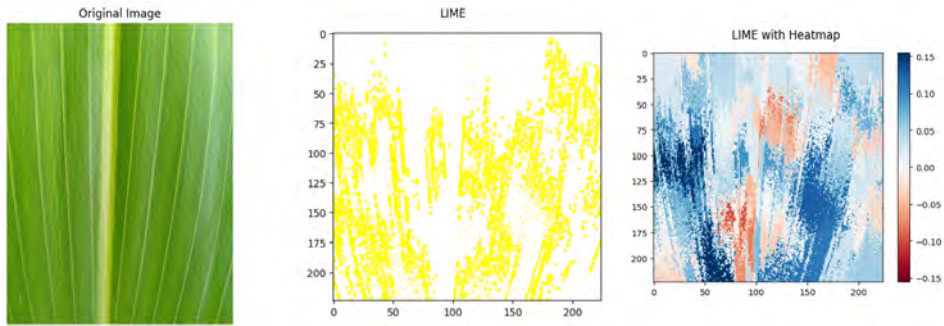


Figure 6.9: LIME results for a health maize crop using fusion model: the original image (left), LIME mask (center), and LIME with heatmap (right).

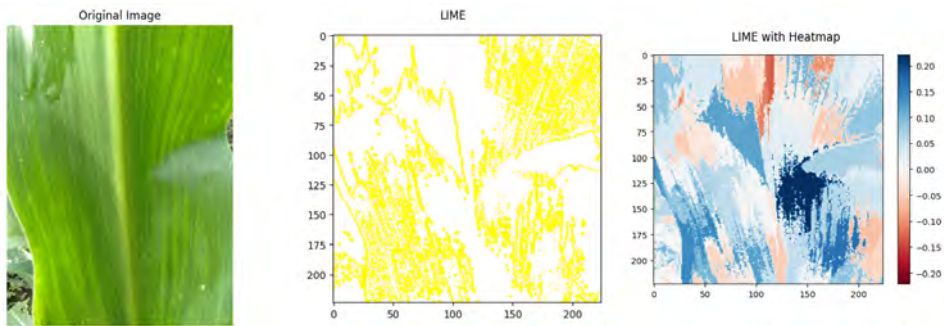


Figure 6.10: LIME results for a mln maize crop using fusion model: the original image (left), LIME mask (center), and LIME with heatmap (right).

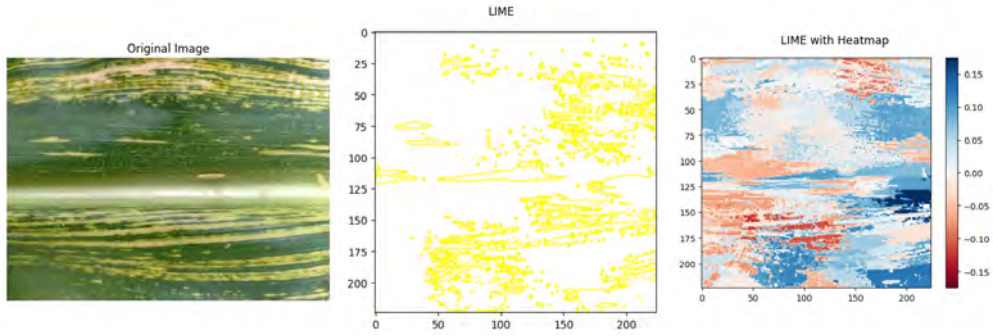


Figure 6.11: LIME results for a msv_1 maize crop using fusion model: the original image (left), LIME mask (center), and LIME with heatmap (right)

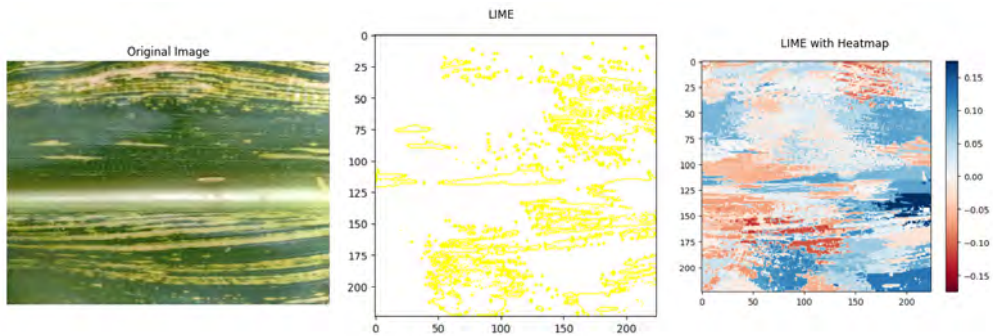


Figure 6.12: LIME results for a msv_2 maize crop using fusion model: the original image (left), LIME mask (center), and LIME with heatmap (right).

To summarize, the diverse XAI techniques employed in our maize disease detection study, namely saliency maps, Grad-CAM, and LIME, each play a distinct role in enhancing the model's transparency and interpretability. The saliency map shows users how the model prioritizes certain features, such as colors and textures that indicate disease symptoms. Grad-CAM makes predictions that let users see how the model prioritizes certain features, such as colors and textures that show disease symptoms. LIME further complements these methods by breaking down the image into superpixels, helping to isolate and explain how each segment influences the model's output. Together, these techniques create a comprehensive framework for interpreting complex deep learning models, ensuring that the model's decision-making processes are not only transparent but also aligned with domain expertise, ultimately validating the model's reliability and accuracy in practical applications.

Chapter 7

Conclusion and Future Direction

7.1 Conclusion

This research has demonstrated the powerful impact of advanced deep learning models, particularly Vision Transformers, combined with Explainable AI (XAI) techniques, on improving maize disease detection. By analyzing leading Transfer Learning models (EfficientNetV2B2, ResNet50, InceptionV3, VGG16, and Xception) alongside Vision Transformer architectures (SWIN, DaViT, MobileViT, MaxViT, and Involucional Neural Networks), we identified the fusion of MobileViT and DaViT as the top performer, achieving a remarkable diagnostic accuracy of 96.67. This model not only outperformed individual approaches but also proved to be practical, with successful deployment on a Raspberry Pi, providing a scalable solution for farmers in resource-limited environments. The integration of XAI techniques, such as Grad-CAM, LIME, and Saliency Maps, further enhanced the model by making its predictions transparent and easy to interpret, enabling farmers to trust and adopt this technology without needing deep technical knowledge. Overall, this research not only achieved state-of-the-art results but also addressed the practical needs of farmers by offering a cost-effective, interpretable, and scalable solution for early disease detection. It lays a strong foundation for future developments in agricultural AI, with the potential to significantly enhance food security and improve sustainable farming practices globally.

7.2 Future Direction

The success of this research opens promising avenues for advancing maize disease detection and expanding its impact. Building on the strong performance of our fusion model, future work could involve expanding the dataset to include a wider range of maize diseases, increasing the model's applicability across diverse agricultural settings. Further optimization of the model could unlock even higher performance, positioning it as a leading solution for crop disease management. Integrating IoT technologies for real-time monitoring would offer farmers a proactive tool for crop protection. The explainability features already make the model accessible to non-experts, and further innovations in this area will drive greater trust and adoption. These advancements have the potential to significantly enhance food security, improve agricultural productivity, and empower farmers with advanced technology.

Bibliography

- [1] B. Shiferaw, B. Prasanna, J. Hellin, and M. Bänziger, “Crops that feed the world 6. past successes and future challenges to the role played by maize in global food security,” *Food Security*, vol. 3, no. 3, pp. 307–327, 2011. DOI: 10.1007/s12571-011-0140-5. [Online]. Available: <https://doi.org/10.1007/s12571-011-0140-5>.
- [2] O. Russakovsky, J. Deng, H. Su, *et al.*, “Imagenet large scale visual recognition challenge,” *CoRR*, vol. abs/1409.0575, 2014. arXiv: 1409.0575. [Online]. Available: <http://arxiv.org/abs/1409.0575>.
- [3] K. Simonyan and A. Zisserman, “Very deep convolutional networks for Large-Scale image recognition,” *arXiv (Cornell University)*, Jan. 2014. DOI: 10.48550/arxiv.1409.1556. [Online]. Available: <https://arxiv.org/abs/1409.1556>.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *arXiv (Cornell University)*, Jan. 2015. DOI: 10.48550/arxiv.1512.03385. [Online]. Available: <https://arxiv.org/abs/1512.03385>.
- [5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” *arXiv (Cornell University)*, Jan. 2015. DOI: 10.48550/arxiv.1512.00567. [Online]. Available: <https://arxiv.org/abs/1512.00567>.
- [6] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” *arXiv (Cornell University)*, Jan. 2016. DOI: 10.48550/arxiv.1610.02357. [Online]. Available: <https://arxiv.org/abs/1610.02357>.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
- [8] W. Nash, T. Drummond, and N. Birbilis, “A review of deep learning in the study of materials degradation,” *npj materials degradation*, vol. 2, no. 1, Nov. 2018. DOI: 10.1038/s41529-018-0058-x. [Online]. Available: <https://doi.org/10.1038/s41529-018-0058-x>.
- [9] N. Parmar, A. Vaswani, J. Uszkoreit, *et al.*, “Image Transformer,” *arXiv (Cornell University)*, Jan. 2018. DOI: 10.48550/arxiv.1802.05751. [Online]. Available: <https://arxiv.org/abs/1802.05751>.
- [10] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” *CoRR*, vol. abs/1905.11946, 2019. arXiv: 1905.11946. [Online]. Available: <http://arxiv.org/abs/1905.11946>.

- [11] Y.-Y. Zheng, J.-L. Kong, X.-B. Jin, X.-Y. Wang, T.-L. Su, and M. Zuo, “Cropdeep: The crop vision dataset for deep-learning-based classification and detection in precision agriculture,” *Sensors*, vol. 19, no. 5, p. 1058, 2019. DOI: 10.3390/s19051058. [Online]. Available: <https://doi.org/10.3390/s19051058>.
- [12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *arXiv (Cornell University)*, Jan. 2020. DOI: 10.48550/arxiv.2010.11929. [Online]. Available: <https://arxiv.org/abs/2010.11929>.
- [13] K. P. Panigrahi, H. Das, A. K. Sahoo, and S. C. Moharana, “Maize leaf disease detection and classification using machine learning algorithms,” in *Progress in Computing, Analytics and Networking*, H. Das, P. K. Pattnaik, S. S. Rautaray, and K.-C. Li, Eds., Singapore: Springer Singapore, 2020, pp. 659–669, ISBN: 978-981-15-2414-1.
- [14] Z. Liu, Y. Lin, Y. Cao, *et al.*, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” *arXiv (Cornell University)*, Jan. 2021. DOI: 10.48550/arxiv.2103.14030. [Online]. Available: <https://arxiv.org/abs/2103.14030>.
- [15] S. Mehta and M. Rastegari, *MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer*, Oct. 2021. [Online]. Available: <https://arxiv.org/abs/2110.02178>.
- [16] M. Tan and Q. V. Le, “EfficientNetV2: Smaller models and faster training,” *arXiv (Cornell University)*, Jan. 2021. DOI: 10.48550/arxiv.2104.00298. [Online]. Available: <https://arxiv.org/abs/2104.00298>.
- [17] J. Van Der Putten and F. G. Zanjani, *Multi-scale ensemble of ResNet variants*. Jan. 2021, pp. 115–119. DOI: 10.1007/978-3-030-64340-9\{-}13. [Online]. Available: https://doi.org/10.1007/978-3-030-64340-9_13.
- [18] M. Ding, B. Xiao, N. Codella, P. Luo, J. Wang, and L. Yuan, *DAViT: Dual Attention Vision Transformers*, Apr. 2022. [Online]. Available: <https://arxiv.org/abs/2204.03645>.
- [19] O. Erenstein, M. Jaleta, K. Sonder, K. Mottaleb, and B. Prasanna, “Global maize production, consumption and trade: Trends and r&d implications,” *Food Security*, vol. 14, no. 14, 2022. DOI: 10.1007/s12571-022-01288-7. [Online]. Available: <https://doi.org/10.1007/s12571-022-01288-7>.
- [20] M. A. Haque, S. Marwaha, C. K. Deb, *et al.*, “Deep learning-based approach for identification of diseases of maize crop,” *Scientific Reports*, vol. 12, p. 6334, 2022. DOI: 10.1038/s41598-022-10140-z. [Online]. Available: <https://doi.org/10.1038/s41598-022-10140-z>.
- [21] J. He, T. Liu, L. Li, Y. Hu, and G. Zhou, “Mfaster r-cnn for maize leaf diseases detection based on machine vision,” *Arabian Journal for Science and Engineering*, May 2022. DOI: 10.1007/s13369-022-06851-0. [Online]. Available: <https://doi.org/10.1007/s13369-022-06851-0>.
- [22] R. Khan, M. A. Khan, M. A. Ansari, N. Dhingra, and N. Bhati, *Machine learning-based agriculture*. Jan. 2022, pp. 3–27. DOI: 10.1016/b978-0-323-90550-3.00003-5. [Online]. Available: <https://doi.org/10.1016/b978-0-323-90550-3.00003-5>.

- [23] N. Kundu, G. Rani, V. S. Dhaka, *et al.*, “Disease detection, severity prediction, and crop loss estimation in maizecrop using deep learning,” *Artificial Intelligence in Agriculture*, vol. 6, pp. 276–291, 2022, ISSN: 2589-7217. DOI: <https://doi.org/10.1016/j.aiaa.2022.11.002>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2589721722000204>.
- [24] Y. Li, S. Sun, C. Zhang, G. Yang, and Q. Ye, “One-stage disease detection method for maize leaf based on multi-scale feature fusion,” *Applied Sciences*, vol. 12, no. 16, 2022, ISSN: 2076-3417. DOI: 10.3390/app12167960. [Online]. Available: <https://www.mdpi.com/2076-3417/12/16/7960>.
- [25] Z. Li, G. Zhou, Y. Hu, *et al.*, “Maize leaf disease identification based on wgmarnet,” *PLOS ONE*, vol. 17, no. 4, e0267650, 2022. DOI: 10.1371/journal.pone.0267650. [Online]. Available: <https://doi.org/10.1371/journal.pone.0267650>.
- [26] H. Liu, H. Lv, J. Li, Y. Liu, and L. Deng, “Research on maize disease identification methods in complex environments based on cascade networks and two-stage transfer learning,” *Scientific Reports*, vol. 12, no. 1, 2022. DOI: 10.1038/s41598-022-23484-3. [Online]. Available: <https://doi.org/10.1038/s41598-022-23484-3>.
- [27] J. A. L. Marques, F. N. B. Gois, J. P. D. V. Madeiro, T. Li, and S. J. Fong, *Artificial neural network-based approaches for computer-aided disease diagnosis and treatment*. Jan. 2022, pp. 79–99. DOI: 10.1016/b978-0-323-85751-2.00008-6. [Online]. Available: <https://doi.org/10.1016/b978-0-323-85751-2.00008-6>.
- [28] E. F. Mohammad Fraiwan and N. Khasawneh, “Classification of corn diseases from leaf images using deep transfer learning,” *Plants (Basel, Switzerland)*, vol. 11, no. 20, p. 2668, 2022. DOI: 10.3390/plants11202668. [Online]. Available: <https://doi.org/10.3390/plants11202668>.
- [29] A. S. Paymode and V. B. Malode, “Transfer learning for multi-crop leaf disease image classification using convolutional neural network vgg,” *Artificial Intelligence in Agriculture*, vol. 6, pp. 23–33, 2022, ISSN: 2589-7217. DOI: <https://doi.org/10.1016/j.aiaa.2021.12.002>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2589721721000416>.
- [30] D. Reiff, *Understand your algorithm with grad-cam*, *Medium*, May 2022. [Online]. Available: <https://towardsdatascience.com/understand-your-algorithm-with-grad-cam-d3b62fce353>.
- [31] D. Sutaji and O. Yıldız, “Lemoxinet: Lite ensemble mobilenetv2 and xception models to predict plant disease,” *Ecological Informatics*, vol. 70, p. 101698, 2022, ISSN: 1574-9541. DOI: <https://doi.org/10.1016/j.ecoinf.2022.101698>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1574954122001480>.
- [32] Z. Tu, H. Talebi, H. Zhang, *et al.*, *MaxViT: Multi-Axis Vision Transformer*, Apr. 2022. [Online]. Available: <https://arxiv.org/abs/2204.01697>.
- [33] S. Ali, T. Abuhmed, S. El-Sappagh, *et al.*, “Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence,” *Information fusion*, vol. 99, p. 101805, 2023.

- [34] M. Banadda, N. K. Aloysius, S. Nakazzi, O. B. Ernest, M. S. Owekitiibwa, and G. Marvin, “Explainable artificial intelligence for maize disease diagnostics,” in *2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 2023, pp. 1–6. DOI: 10.1109/CSDE59766.2023.10487660.
- [35] A. Dash, P. K. Sethy, and S. K. Behera, “Maize disease identification based on optimized support vector machine using deep feature of densenet201,” *Journal of Agriculture and Food Research*, vol. 14, p. 100 824, 2023, ISSN: 2666-1543. DOI: <https://doi.org/10.1016/j.jafr.2023.100824>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666154323003319>.
- [36] Y. Hu, G. Liu, Z. Chen, J. Liu, and J. Guo, “Lightweight one-stage maize leaf disease detection model with knowledge distillation,” *Agriculture*, vol. 13, no. 9, 2023, ISSN: 2077-0472. DOI: 10.3390/agriculture13091664. [Online]. Available: <https://www.mdpi.com/2077-0472/13/9/1664>.
- [37] M. M. Islam, M. A. Talukder, M. R. A. Sarker, *et al.*, “A deep learning model for cotton disease prediction using fine-tuning with smart web application in agriculture,” *Intelligent Systems with Applications*, vol. 20, p. 200 278, 2023, ISSN: 2667-3053. DOI: <https://doi.org/10.1016/j.iswa.2023.200278>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2667305323001035>.
- [38] M. Jung, J. S. Song, A.-Y. Shin, *et al.*, “Construction of deep learning-based disease detection model in plants,” *Scientific Reports*, vol. 13, p. 7331, 2023. DOI: 10.1038/s41598-023-34549-2. [Online]. Available: <https://doi.org/10.1038/s41598-023-34549-2>.
- [39] D. M. Kabala, A. Hafiane, L. Bobelin, and R. Canals, “Image-based crop disease detection with federated learning,” *Scientific Reports*, vol. 13, p. 19 220, 2023. DOI: 10.1038/s41598-023-46218-5. [Online]. Available: <https://doi.org/10.1038/s41598-023-46218-5>.
- [40] M. F. Kalyango and K. M. Ntanda, “Interpretable deep learning for diagnosis of maize streak disease,” in *2023 First International Conference on the Advancements of Artificial Intelligence in African Context (AAIAC)*, 2023, pp. 1–6. DOI: 10.1109/AAIAC60008.2023.10465315.
- [41] F. Khan, N. Zafar, M. Naveed, M. A. Tahir, H. Waheed, and Z. Haroon, “A mobile-based system for maize plant leaf disease detection and classification using deep learning,” *Frontiers in Plant Science*, vol. 14, 2023. DOI: 10.3389/fpls.2023.1079366. [Online]. Available: <https://doi.org/10.3389/fpls.2023.1079366>.
- [42] A. MP and P. Reddy, “Ensemble of cnn models for classification of groundnut plant leaf disease detection,” *Smart Agricultural Technology*, vol. 6, p. 100 362, 2023, ISSN: 2772-3755. DOI: <https://doi.org/10.1016/j.atech.2023.100362>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2772375523001909>.
- [43] M. Masood, M. Nawaz, T. Nazir, *et al.*, “Maizenet: A deep learning approach for effective recognition of maize plant leaf diseases,” *IEEE Access*, vol. 11, pp. 52 862–52 876, 2023. DOI: 10.1109/ACCESS.2023.3280260.

- [44] S. D. Meena, M. Susank, T. Guttula, S. H. Chandana, and J. Sheela, “Crop yield improvement with weeds, pest and disease detection,” *Procedia Computer Science*, vol. 218, pp. 2369–2382, 2023, International Conference on Machine Learning and Data Engineering, ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2023.01.212>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050923002120>.
- [45] L. Nazari, M. F. Aslan, K. Sabanci, *et al.*, “Integrated transcriptomic meta-analysis and comparative artificial intelligence models in maize under biotic stress,” *Scientific Reports*, vol. 13, p. 15 899, 2023. DOI: 10.1038/s41598-023-42984-4. [Online]. Available: <https://doi.org/10.1038/s41598-023-42984-4>.
- [46] S. Suara, A. Jha, P. Sinha, and A. A. Sekh, “Is grad-cam explainable in medical images?” In *International Conference on Computer Vision and Image Processing*, Springer, 2023, pp. 124–135.
- [47] H.-T. Vo, L.-D. Quach, and H. T. Ngoc, “Ensemble of deep learning models for multi-plant disease classification in smart farming,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, 2023. DOI: 10.14569/IJACSA.2023.01405108. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2023.01405108>.
- [48] K. Alam, M. F. Mridha, S. Alfarhood, M. Safran, M. Abdullah-Al-Jubair, and D. Che, “A robust and light-weight transfer learning-based architecture for accurate detection of leaf diseases across multiple plants using less amount of images,” *Frontiers in Plant Science*, vol. 14, 2024. DOI: 10.3389/fpls.2023.1321877. [Online]. Available: <https://doi.org/10.3389/fpls.2023.1321877>.
- [49] M. Alkanan and Y. Gulzar, “Enhanced corn seed disease classification: Leveraging mobilenetv2 with feature augmentation and transfer learning,” *Frontiers in Applied Mathematics and Statistics*, vol. 9, 2024. DOI: 10.3389/fams.2023.1320177. [Online]. Available: <https://doi.org/10.3389/fams.2023.1320177>.
- [50] I. Khan, S. S. Sohail, D. Ø. Madsen, and B. K. Khare, “Deep transfer learning for fine-grained maize leaf disease classification,” *Journal of Agriculture and Food Research*, vol. 16, p. 101 148, 2024, ISSN: 2666-1543. DOI: <https://doi.org/10.1016/j.jafr.2024.101148>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666154324001856>.
- [51] T. J. Maginga, E. Masabo, P. Bakunzibake, K. S. Kim, and J. Nsenga, “Using wavelet transform and hybrid cnn – lstm models on voc ultrasound iot sensor data for non-visual maize disease detection,” *Heliyon*, vol. 10, no. 4, e26647, 2024. DOI: 10.1016/j.heliyon.2024.e26647. [Online]. Available: <https://doi.org/10.1016/j.heliyon.2024.e26647>.
- [52] A. S. Md. Siam, A. Hossain, R. B. Hossain, and M. M. Rahman, “Se-vgg16 maizenet: Maize disease classification using deep learning and squeeze and excitation attention networks,” in *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2024, pp. 1–6. DOI: 10.1109/ESCI59607.2024.10497322.

- [53] T. O'Halloran, G. Obaido, B. Otegbade, and I. D. Mienye, "A deep learning approach for maize lethal necrosis and maize streak virus disease detection," *Machine Learning with Applications*, vol. 16, p. 100 556, 2024, ISSN: 2666-8270. DOI: <https://doi.org/10.1016/j.mlwa.2024.100556>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S266682702400032X>.
- [54] M. H. Al-Qadi, M. F. El-Habibi, R. Z. Sababah, and S. S. Abu-Naser, "Using deep learning to classify corn diseases," *International Journal of Academic Information Systems Research (IJAIRS)*, vol. 8, no. 4, pp. 81–88, 2024, ISSN: 2643-9026. [Online]. Available: <http://ijeais.org/wp-content/uploads/2024/4/IJAISR240411.pdf>.
- [55] S. T. Y. Ramadan, T. Sakib, R. Jahangir, and S. Rahman, "Maize leaf disease detection using vision transformers (vits) and cnn-based classifiers: Comparative analysis," in *Human-Centric Smart Computing*, S. Bhattacharyya, J. S. Banerjee, and M. Köppen, Eds., Singapore: Springer Nature Singapore, 2024, pp. 513–524, ISBN: 978-981-99-7711-6.
- [56] Statista, *Production volume of maize in Africa 2017-2025*, Oct. 2024. [Online]. Available: <https://www.statista.com/statistics/1294303/production-volume-of-corn-in-africa/#statisticContainer>.
- [57] M. Wambui, "Identification of maize leaf diseases based on alexnet and resnet50 convolutional neural networks," *Indonesian Journal of Computer Science*, 2024, Retrieved May 20, 2024, from https://www.academia.edu/113827224/Identification_of_Maize_Leaf_Diseases_Based_On_AlexNet_and_ResNet50_Convolutional_Neural_Networks. [Online]. Available: https://www.academia.edu/113827224/Identification_of_Maize_Leaf_Diseases_Based_On_AlexNet_and_ResNet50_Convolutional_Neural_Networks.
- [58] H. Wu, "Maize leaf disease image classification based on resnet18 image classification model," in *Proceedings Volume 13105, International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2023)*, (2023), Mar. 2024, p. 131050V. DOI: 10.1117/12.3026727. [Online]. Available: <https://doi.org/10.1117/12.3026727>.