Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Agentic AI Based Plant Disease Detection System

J Ramprasath
Dr. Mahalingam College of Engineering and Technology
Deekshiba PV
Dr. Mahalingam College of Engineering and Technology
Jaya Rupa A A
Dr. Mahalingam College of Engineering and Technology
Ashef S
Dr. Mahalingam College of Engineering and Technology
Sheikh Shafin Ahmad
The University of Texas at El Paso
Somenath Chakraborty
West Virginia University Institute of Technology
Saurav Mallik
smallik@utep.edu

The University of Texas at El Paso

Research Article

Keywords: Agentic Artificial Intelligence, Plant Disease Detection, MobileNetV2, Deep Learning, Precision Agriculture, Edge Deployment

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11671402/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Agentic AI Based Plant Disease Detection System

J Ramprasath[1,*], Deekshiba PV[2], Jaya Rupa A A[2], Ashef S[2], Sheikh Shafin Ahmad[3], Somenath Chakraborty[4], Saurav Mallik[5,6,*]

[1]Department of Information Technology, Dr. Mahalingam College of Engineering and Technology, Pollachi, India; jrprasath@gmail.com

[2]Department of Artificial Intelligence and Data Science, Dr. Mahalingam College of Engineering and Technology, Pollachi, India; deekshibapv@gmail.com, jayarupa09@gmail.com, hafizashef@gmail.com

[3]Department of Environmental Science and Engineering, University of Texas at El Paso, El Paso, TX 79968, USA; ssahmad@miners.utep.edu

[4]Department of Computer Science and Information Systems, The West Virginia University Institute of Technology, Beckley, West Virginia, USA; somenath.chakraborty@mail.wvu.edu

[5]Department of Environmental Health, Harvard T H Chan School of Public Health, Boston, MA, USA; sauravmtech2@gmail.com, smallik@hsph.harvard.edu

[6]Department of Biological Sciences, University of Texas at El Paso, El Paso, TX 79968, USA; smallik@utep.edu

***Correspondence:** J Ramprasath (jrprasath@gmail.com); Saurav Mallik (sauravmtech2@gmail.com, smallik@utep.edu)

Abstract

Problems with plant diseases put global food supply at risk; they result in many crops not being able to produce as much or in lower quality than would otherwise be produced. Many small-scale farmers have difficulty accurately and in a timely manner diagnosing disease because of the high expense and limited availability of traditional inspection methods. This study suggests using a novel framework based on agentic artificial intelligence (AI), which incorporates deep learning-based visual analysis and the use of relevant context within the environment, to achieve new levels of intelligent crop health evaluation. The proposed solution is a collaborative multi-agent architecture where dedicated agents can independently complete all steps: automated image preprocessing, automated disease classification, environmental risk assessment, and treatment recommendations. The study employs MobileNetV2 as its backbone because of its lightweight design and low computational requirements, making it possible to use within the resource-limited agriculture environment. To further enhance the quality of diagnostic testing over the long term, there is an added adaptive learning component involved

that incorporates feedback from the previous learning cycles. The PlantVillage dataset was used for an empirical analysis and showed in the proposed methodology has a total overall accuracy of 95.32% for the classification. The macro-AVG. for precision 0.95, recall 0.95 and F1 Score 0.95 are indicate the ability to predict well, with stable convergence. The system performs computations with less energy than other systems and can be deployed in a real-time manner on small devices. The transition from intelligent decision support systems to disease diagnostics will be an important milestone toward evolving scalable and precise agricultural systems as well as future sustainable crop production.

Keywords: *Agentic Artificial Intelligence; Plant Disease Detection; MobileNetV2; Deep Learning; Precision Agriculture; Edge Deployment*.

1. Introduction

Agricultural production is essential to providing the world with a stable food supply and economic prosperity, especially in the developing world, where many people rely on agriculture as a source of income. However, agricultural production is increasingly threatened by the effects of plant diseases that reduce yield and quality. (CABI, 2023). Plant diseases negatively impact agricultural production by interfering with both the quantity or quality of harvests, causing disruptions to the supply chain and creating an additional burden for the global economy. Therefore, detecting plant diseases quickly and accurately is essential for establishing sustainable agricultural practices.

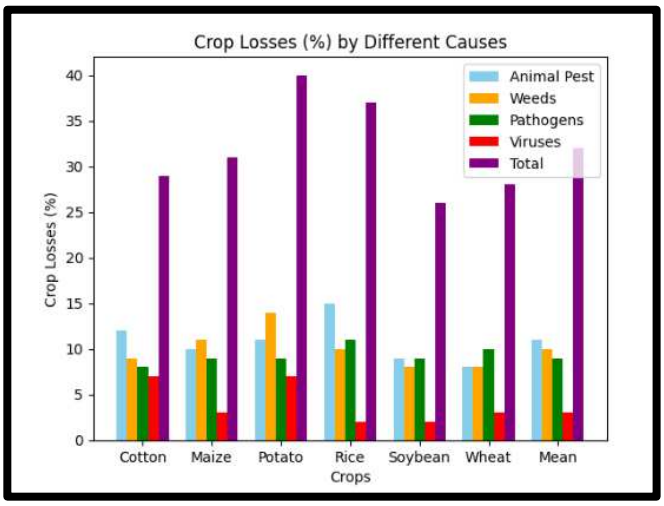


Figure 1: Percentage Crop Yield Losses Caused by Pests, Weeds, Pathogens, and Viruses Across Major Agricultural Crops.

One of the main sources of food insecurity globally is through plant diseases. The FAO estimates that over 40% of annual food production is lost globally to pests and plants diseases and that pest and disease-related trade losses exceed $220 billion. Invasive species and their

associated pests are currently costing global economies almost $70 billion each year, and food production will need to increase by approximately 70% by 2050 to accommodate anticipated populations over this period. This information indicates that protecting crop health is becoming more important every day. Additionally, since about 80% of all the food consumed worldwide comes from plants, decreased plant health will continue to have adverse effects on nutritional security, income for farmers, and stability of ecosystems.

Crop loss percentages in **Figure 1** show the contribution animals, weeds, disease pathogens, and viruses make to the crop that are typically 25% to 40% of total crops lost. Plant diseases and viruses cause the largest share of crop losses. Losses also increase the severity of food insecurity experienced by millions globally. Protecting crops from getting diseased is an urgent agricultural and a humanitarian need.

Plant disease diagnosis methods today involve highly trained ag experts visually inspecting crops for diseases. While these methods give accurate results, they are very labour intensive, take a long time to complete, and are inaccessible to most small farmers, particularly rural farmers with little to no access to ag experts. The delays in diagnosing crop diseases will lead to incorrectly treating diseased crops and cause additional crop loss. Therefore, there is a need for intelligent and automated and low-cost diagnostic systems to diagnosis crops in resource-constrained environments.

The new advancements in AI and deep learning extremely affect the identifying plant disease from images. Convolutional neural networks (CNN) are used to identify plant disease through image analysis are more accurately than any other method. Various network architectures (e.g., VGG16, ResNet50, EfficientNet, MobileNet) have been tested for their potential to reliably improve predictions and provide a more efficient computationally complex way of doing the same.

Larger and deeper networks, such as ResNet50, tend to achieve greater overall accuracy because they can capture more intricate hierarchical representations of the data. Unfortunately, these larger networks will not be able to support live operations due to their computation costs. Smaller networks (like MobileNetV2) have been found to be the best candidates in terms of providing an appropriate balance between performance and power requirements. These networks can be effectively utilized on low power, limited storage devices typically found at the edge of networks. Despite these developments, most current plant disease detection systems operate independently as classifiers that can identify plant diseases. As a result, there is little

context for farmers about the conditions that surrounded the plants or the associated recommendations for farmers looking to address the conditions of their plants.

Plant disease expression is highly affected by temperature, moisture, and seasons; environmental influences may alter how we view a plant. A system that depends solely upon visual characteristics is unlikely to include the influence of environmental factors and therefore provide less reliable diagnoses under real-world agricultural conditions. This limitation creates a need for intelligent systems that combine the use of visual perception and environmental perception to facilitate effective agricultural decision-making processes.

There is still considerable research required to fill the gaps in the AI space concerning plant health monitoring. Much of the work currently done is based on individual datasets and as such is limited in the generalization of the model when applied to various agricultural contexts. Further, little has been done to formally compare deep learning architectures to measure their robustness, efficiency, and adaptability in various environmental conditions. Developing a robust but efficient framework capable of transforming diverse data into valid outcomes will establish an intelligent method of addressing the difficulty associated with meeting the challenges of making a valid decision regarding the suitability of a given crop to be grown in a specific location based upon a wide range of data types will require a framework that can be scaled up and made to run efficiently on relatively less powerful computing systems [8, 14, 16].

The AAI will achieve this by using a combination of an AAI structure that will allow the development of a low-complexity, multi-agent collaborative framework of independent agents executing multiple ways in which each agent will be responsible for executing key functions associated with the completion of various processes related to the identification of plant diseases and the management of crop health, including activities such as image analysis, the classification of plant disease incidences, risk evaluation of the environment, acquiring knowledge and making recommendations for the treatment of diseased plants. The distributed architecture allows the AAI to provide increased flexibility and scalability for the decision-making capability of the users [18].

With a small footprint, few parameters, and high accuracy with low computational cost, MobileNetV2 was selected as the backbone network for agricultural environments with limited resources. Additionally, an adaptive learning strategy with dynamic feedback allows for continued improvement of model performance over time [10, 11].

To assess the degree of robustness and generalization of the proposed model across domains, an experimental comparison will be made between the proposed model and several well-established architectures (e.g. VGG16 and ResNet50) using multiple publicly available datasets (PlantVillage, FieldPlant, and New Plant Disease dataset). Using these multiple datasets provides a wider range of testing for model reliability at various levels of imaging quality [6, 7, 9].

The purpose of this project is to change how plant diseases are identified from simply identifying how/when to classify the plant disease, to developing intelligent decision support systems that can provide farmers with information close to the real-time planting decision processes. Furthermore, the integration of deep learning with agent-based coordination and awareness of environmental context has the capacity to support the further development of precision agriculture as well as sustainable agricultural practices engaging established best practices.

The proposed Contribution includes the following: Project proposal details provide new developments about multi-agent architecture for disease identification through the provision of decision support for use by the intelligent agriculture industry. Additionally, an alternative method of providing a hybrid diagnosing process using both visual analysis of the leaf and corresponding environmental conditions further increase the reliability and strength for predicting plant diseases. This research will yield a MobileNetV2 model which will be a lightweight version of the model designed to meet the needs of real time applications in resource constrained environments. Furthermore, an adaptable mechanism based upon user providing feedback will be utilised to assist with the ongoing adaptation of the system and enhance ongoing performance of the system above and beyond initial capability. Finally, an extensive analysis of MobileNetV2 when compared to VGG16 and ResNet50 is presented for both prediction and technical efficiency which creates an innovative, new process for real time diagnosing of a plant disease.

The structure of the this paper is: literature related to AI and plant disease diagnosis will be reviewed in **Section 2**; proposed research methodology will be discussed in **Section 3**; experimental design will be detailed in **Section 4**; results of the current study, including performance assessments of those results will be highlighted in **Section 5**; and conclusions from the investigation and future direction of research will be presented in **Section 6**.

2. Related Work

The use of modern imaging techniques combined with artificial intelligence (AI) technology is making automated plant disease diagnosis a reality. While manual feature selection (thereby specifying colour, texture, shape etc.) and traditional ML algorithms like K-Nearest Neighbours (KNN), Random Forests, and Support Vector Machines (SVMs) were used in early disease diagnostic systems, these methods were limited in terms of how well they would perform under different climatic conditions (not sufficiently robust and/or scalable). In contrast, CNNs learn automatically from image data hierarchically to progress the classification accuracy without the need to have created the features manually.

Plant leaf diseases are predicated by Deep CNNs, for example, network structures such as VGG16, VGG19, ResNet and InceptionV3 have produced significant successes in relation to the accuracy of classification when trained on large, well-annotated datasets such as PlantVillage. Moreover, the rapid development of model optimization through model compression techniques and mobile cloud technologies such as TensorFlow Lite are allowing for real-time recognition of diseases using smartphone applications making it even easier for farmers to access this new technology and improve their farming outcomes.

Even though there are some successful examples, laboratory-derived datasets tend not to accurately reflect the complexity of real-world agricultural environments. For example, when your dataset consists of a controlled background, consistent illumination, and minimal obstruction, you may inadvertently overestimate how accurately your model will perform in the field. To help overcome this limitation, many research efforts have developed field-level datasets such as FieldPlant, which consist of images that are taken from farms in natural environmental conditions and using camera techniques. While the findings from benchmarking experiments using MobileNet, VGG16, and InceptionV3 architectures have shown promise, there remain several additional challenges including overlapping leaves; background noise; and variability in the illumination which prevent the generalization of the model results.

Recently, machine-learning researchers have turned to Explainable Artificial Intelligence (XAI) methods to better understand how ML models can boost to the predict of diseases in agricultural diagnostics. ML model explains how various features of a leaf image impacted predictions of disease, it enhances trust and transparency in how decisions are made. In addition, classification rates exceeding 98% for high-performance models such as ResNet50 suggest that deep learning can be a reliable and valuable tool for detecting agricultural diseases.

The hybrid deep learning models have started to attract attention from researchers, consolidating multi-scale feature extraction and enhancing diagnostic accuracy for diseases. For example, the S-ZFFTNet model employs a multi-step approach of using anisotropic diffusion for denoising, employing conditional Generative Adversarial Networks (cGAN) for segmenting leaves and applying various advanced techniques for feature selection, including Fuzzy Local Binary Patterns (FLBP) and wavelet transforms. Table 1 depicts comparative study of Existing Plant Disease Detection Systems.

Table 1: Comparative study of Existing Plant Disease Detection Systems

Study	Technique	Key Contribution	Research Gap	Relevance to Proposed Work
Demilie et al. (2024) [1]	ML and DL comparison focusing on CNN models.	Reviews recent methods and supports multi-class detection.	Relies mainly on lab images, limiting field adaptability.	Highlighted the need for real-world disease detection.
Waghmare et al. (2024) [2]	ResNet50 with environmental features.	Achieves high accuracy while considering climate factors.	Requires large datasets and may overfit.	Encouraged building a data-efficient model.
Kamei (2023) [3]	DetectoRS and YOLOX for resolution analysis.	Maintains performance across different image qualities.	Tested mostly in controlled settings.	Stressed the importance of field validation.
Nigar et al. (2024) [15]	Explainable AI using EfficientNetB0 and LIME.	Provides accurate and interpretable predictions.	Limited dataset diversity affects robustness.	Motivated combining accuracy with transparency.
Moupojou et al. (2023) [4]	Field Plant with original-attribute images.	Introduces an expert marked cassava dataset.	Complex backgrounds reduce detection accuracy.	Supported training with realistic agricultural data.
Nazki et al. (2019) [17]	AR-GAN for synthetic data augmentation.	Improves accuracy by addressing class imbalance.	Sensitive to parameter tuning.	Suggested careful dataset balancing strategies.
Nguyen et al. (2021) [12]	3D-CNN with hyperspectral imaging.	Detects diseases before visible symptoms appear.	Expensive equipment limits practical use.	Reinforced the goal of developing a low-cost system.
Yuan et al. (2025) [13]	FCNN and LSTM for weather-based forecasting.	Predicts disease outbreaks in advance.	Class imbalance may cause false alerts.	Encouraged integrating prediction with balanced learning.
Tiwari et al. (2023) [5]	Multi-stage CNN using	Provides an easy to use system with recommendations.	Depends heavily on dataset size.	Guided focus on usability with stable performance.

	Python and Keras.			
Shelar et al. (2022) [6]	VGG-19 use on mobile via TensorFlow Lite.	Permits instantaneous data detection on smartphones.	Accuracy still needs improvement.	Inspired enhancing precision without losing speed.

The architecture employing the Zeiler-Fergus Network (ZF-Net) and a Siamese CNN is used to improve the feature discrimination process. The model was able to achieve 92.654% Overall accuracy and 90.866% Overall precision when tested on the data set of healthy and diseased leaves, thus providing an efficient, automatic means of diagnosis using this model.

The research to date focuses primarily on improving the accuracy of classification; however, most currently existing hybrid methods provide only disease names and do not provide any value added by providing contextual information or actionable recommendations relative to the observed condition of a plant. Furthermore, environmental factors such as humidity, temperature, and seasonal changes impact disease manifestation, and these have not been integrated into hybrid deep learning model prediction or analysis. Consequently, there is clearly a demand for intelligent systems that can integrate visual perception, context-based intelligence, and adaptive decision-making.

An agentic AI framework with a collaborative multi-agent architecture and integrated with deep learning is being proposed in this research to help close this gap. The proposed system incorporates not only disease classification but also recommended actions based on environmental analyses and will transition from passive detection to intelligent agricultural decision support systems.

3. Proposed Methodology

The study offers a machine-generated framework that automates disease identification and smart crop management through cooperative efforts among many agents. This framework integrates both deep learning imaging technology with real-time environmental context as well as expert-based suggestions for helping farmers and growers to make informed agricultural, production, plant health decisions.

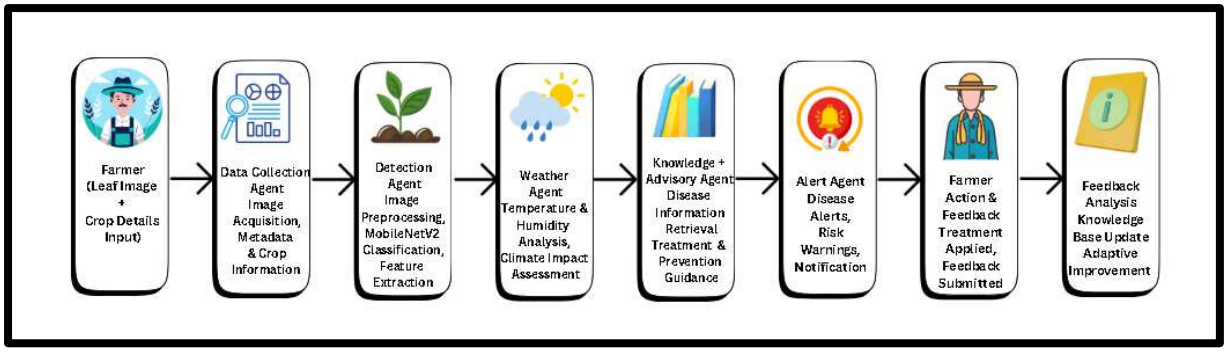


Figure 2: System Architecture of the Proposed Plant Disease Detection System.

In contrast to conventional systems, the described architecture will consist of a multi-agent architecture without a monolithic diagnostic system. The proposed model will include independent agent modules; these will operate sequentially to analyse leaf images, assess risks posed by environmental factors, assess agricultural knowledge, and produce actionable treatment recommendations. **Figure 2** depicts the system architecture of the proposed plant disease detection system.

The overall system can be expressed mathematically as a composite function in equation (1):

$$O = f_{alert}(f_{rec}(f_{env}(f_{cnn}(I)))) \quad \textit{Eq. (1)}$$

where I refer to the input leaf image; f_{cnn} represents the MobileNetV2 based disease detection function; f_{env} is an environmental analysis function; f_{rec} is a knowledge-based recommendation generation function; and f_{alert} provides notifications for farmers.

This composite structure indicates the movement away from passive classifications towards intelligent support for decision-making.

3.1 Dataset Preparation

The PlantVillage dataset has been used for this research. This is a publicly available dataset consisting of 20,638 images of leaves representing 15 types of leaf diseases. To ensure even learning and appropriate validation, the dataset is split 80% for training and 20% for validation as defined in equation (2) and equation (3):

$$D_{train} = 0.8N, D_{val} = 0.2N \quad \textit{Eq. (2)}$$

$$D_{train} = 16{,}516, D_{val} = 4{,}122 \quad \textit{Eq. (3)}$$

As the input size for MobileNetV2 is $224\ x\ 224\ x\ 3$ pixels as defined by the model developer, it has also been resized to match those specifications which is represented in equation (4):

$$I' = Resize(I, 224,224) \qquad Eq.\ (4)$$

The images have also been standardized so they are consistently sized, which helps speed up the computational time required to process them during model training.

3.2 Image Acquisition and Preprocessing

Leaf images obtained from smartphones or digital cameras are sent via web or mobile interface for automated processing. Because agriculture images taken in the field vary greatly regarding illumination, scale, and background complexity, preprocessing is conducted to improve feature characteristics and stabilize model training. **Figure 3** denotes images of leaves: Pre-Processed and Original for Disease Identification.



(a) (b)

Figure 3: Images of Leaves - Pre-Processed and Original for Disease Identification. (a) Original RGB photo of diseased leaf taken in the field, (b) Pre-processed photo of above leaf that has undergone pixel normalisation and contrast enhancements and resized to 224 x 224 to facilitate feature extraction for the MobileNetV2 model. Region of interest is highlighted showing visible symptoms of disease.

i) Resizing the Image

To standardize the images for processing all images are resized to $224\ x\ 224$ pixels by mapping their original coordinates, coordinates (x, y) to a normalized grid:

$$I'(x', y') = I\left(\frac{xW}{224}, \frac{yH}{224}\right) \qquad Eq.\ (5)$$

According to equation (5), the values of W and H are determined from the original pixel dimensions.

ii) Normalizing the Data

MobileNetV2 preprocessing normalizes the pixel intensity values to be the range between $[-1,1]$ as expressed in equation (6):

$$I_n = \frac{I'}{127.5} - 1 \qquad \textit{Eq. (6)}$$

Normalizing data ensures that the distribution of inputs is consistent with that of the weights trained using ImageNet, which helps with stabilizing the gradients and speeding up the time to converge.

iii) Data Augmentation

Stochastic transformations of the image data create multiple versions of the input image; this helps prevent the model from overfitting during training and allows the model to generalize better. This can be mathematically formulated in Equation (7) as:

$$I_{avg} = T(I_n) \qquad \textit{Eq. (7)}$$

T is an augmentation function that contains operations such as rotating, zooming, translating, and flipping an image horizontally.

$$R(\theta) = [cos\ cos\ \theta \quad -sin\ sin\ \theta \quad sin\ sin\ \theta \quad cos\ cos\ \theta\ ]$$

Eq. (8)

The rotation matrix expressed as equation (8) is used for rotating images.

$$D_{eff} = D_{train} + T(D_{train}) \qquad \textit{Eq. (9)}$$

As mentioned in equation (9), the additional transformed images help in creating many more unique images that helps expanding the size of the training set.

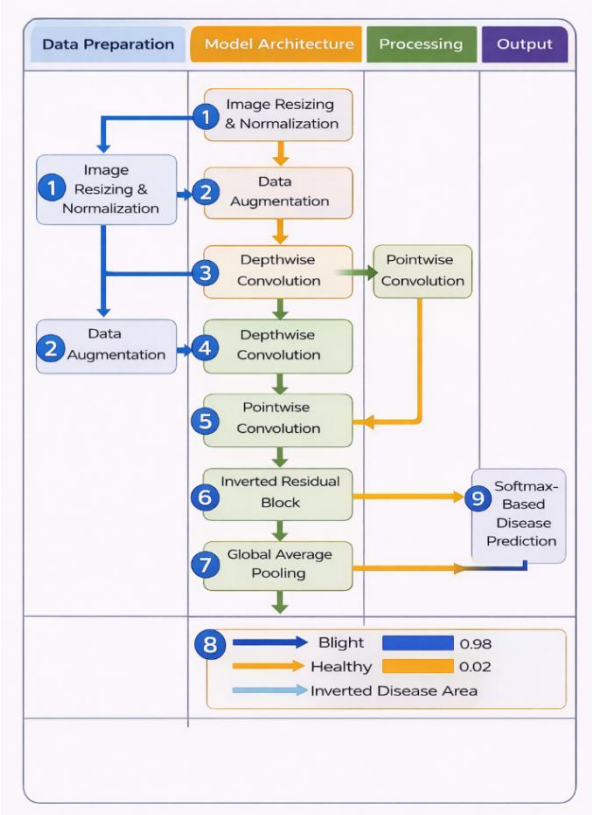


Figure 4: Step Model Based Cross-Functional Flowchart for Potato Early Blight Disease Detection using MobileNetv2.

3.3 MobileNetV2-Based Disease Detection

The decision to use MobileNetV2 primarily stems from its limited number of parameters, small size and its ability to run on an edge device. The overall functionality of the disease identification system is demonstrated in **Figure 4** through a complete workflow of operations. A cross-functional flowchart illustrates the step-by-step series of activities used to capture an image with a camera, preprocess it, extract its features using MobileNetV2, evaluate environmental data then generate recommendations for the user. This is how an agency-based system can be used to perform automated, intelligent assessment of crop health.

3.3.1 Computational Efficiency

Standard convolution has a standard convolution cost defined in equation (10):

$$Cost_{std} = D_k^2 . M . N . D_f^2 \qquad Eq.\ (10)$$

where D_k is kernel size, M input channels, N output channels, and D_f feature map size

The computation with MobileNetV2 will be computed using depth wise separable convolution, whose computational cost is expressed in equation (11):

$$Cost_{dw} = D_k^2 . M . D_f^2 + M . N . D_f^2 \quad \text{Eq. (11)}$$

Using the equation (10) and equation (11), the efficiency of the two methods can be compared to obtain reduction in computation as formulated in equation (12):

$$Reduction \approx \frac{1}{N} + \frac{1}{D_k^2} \quad \text{Eq. (12)}$$

This efficiency is especially useful when using MobileNetV2 for agricultural applications in real-time.

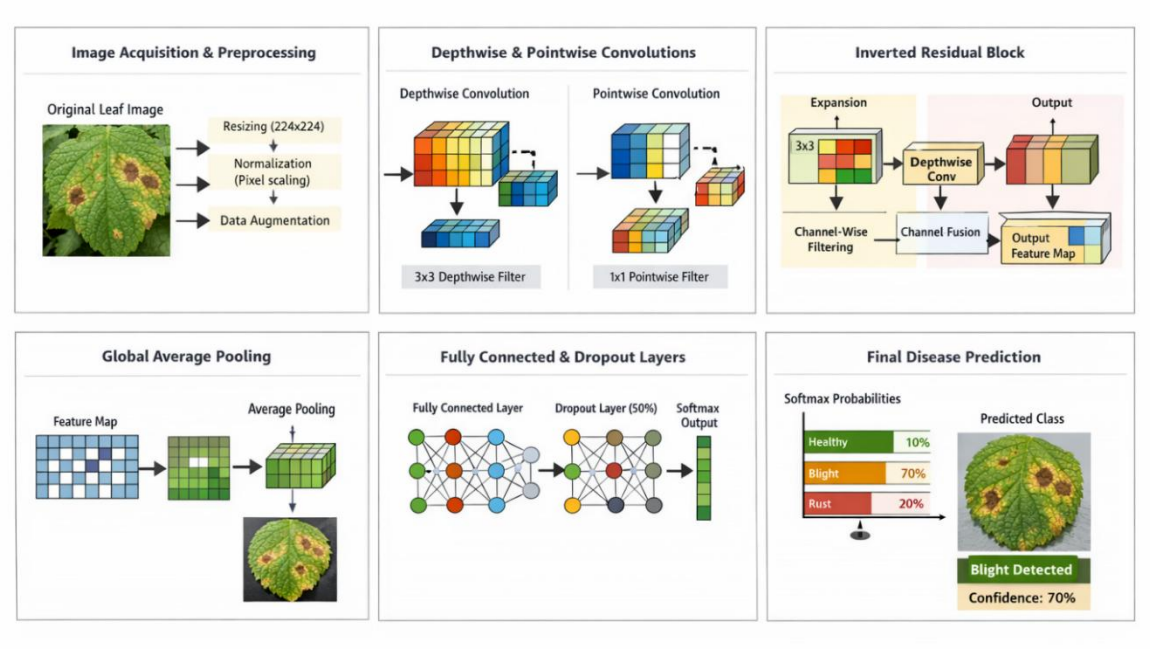


Figure 5: Architecture and Processing Pipeline of MobileNetV2 for Potato Leaf Disease Classification.

As shown in **Figure 5**, MobileNetV2's internal architecture was used to classify diseases. This model employs depth wise separable convolutions to lower computation costs and maintain rich features within the model. This associated pipeline demonstrates the different stages (i.e., filtering in space, fusing channels, applying nonlinear activation, and applying residual learning) which aid in efficiently extracting patterns that are specific to diseases based upon leaf images.

3.3.2 Feature Extraction

Channel-wise extraction of spatial features is accomplished by applying depthwise convolution, as defined in equation (13):

$$Z_k = X_k * K_k \quad \text{Eq. (13)}$$

where X_k represents the k^{th} channel and K_k is its associated kernel.

Cross-channel fusion is accomplished by utilizing pointwise convolution, expressed in equation (14):

$$Y_j = \sum_{k=1}^{M} Z_k W_{k,j} \quad \textit{Eq. (14)}$$

allowing the network to learn distinguishing characteristics of disease, e.g. lesion characteristics, pigmentation, or chlorosis.

3.3.3 Inverted Residual Blocks

The Inverted Residual Block with Linear Bottlenecks significantly improves MobileNetV2 because it first elongates its input feature maps into higher dimensions. The process of applying depth wise convolution allows all the feature sets to be projected onto lower dimensional spaces with a linear activation function while retaining their spatial information. In addition to allowing this approximation of the input features, using linear bottlenecks prohibits the potential for information loss resulting from the use of non-linear activation functions when projecting the input features into lower dimensions. Because of these features, complex pictorial patterns such as:

i. Irregular brown lesions
ii. Areas of diseases
iii. Differences in colour and texture of leaves

can now be learned much more easily. Low-dimensional features of X are expanded through MobileNetV2, as defined in equation (15):

$$X_{exp} = ReLU6(W_e X) \quad \textit{Eq. (15)}$$

Spatial dependencies are captured by depth wise convolution, expressed in equation (16):

$$X_{dw} = ReLU6(K_{dw} * X_{exp}) \quad \textit{Eq. (16)}$$

Representations are compressed through linear projection, as given in equation (17):

$$X_{proj} = W_p X_{dw} \quad \textit{Eq. (17)}$$

Residual connections will keep information intact, as shown in equation (18):

$$Y = X + X_{proj} \quad \textit{Eq. (18)}$$

This structure supports fine-grained disease detection without excessive memory overhead.

3.3.4 Nonlinear Activation (ReLU6)

MobileNetV2 employs ReLU6 activation function which is defined in equation (19):

$$ReLU6(x) = min(max(0, x),6) \quad \textit{Eq. (19)}$$

Bounding activations improves numerical stability on mobile processors and prevents exploding activations.

3.3.5 Global Average Pooling

Using global average pooling as opposed to dense spatial layers condenses the feature maps into a single feature vector for each channel although it does not combine features based on spatial dimensions thus still preserving channel characteristics with fewer parameters which would help reduce the likelihood of overfitting that is expressed in equation (20)

$$G_c = \frac{1}{H\,x\,W}\sum_{i=1}^{H} \quad \sum_{j=1}^{W} \quad F_c(i,j) \quad \textit{Eq. (20)}$$

Once this pool of features has been created for each of the channels it is transformed by a fully connected layer which is formulated in equation (21).

$$h = ReLU(W_h G + b_h) \quad \textit{Eq. (21)}$$

where $W_h \in R^{256xn}$.

To add another layer of regularization, training dropout is used as defined in equation (22)

$$h' = h\,.Bernoulli(p), p = 0.5 \quad \textit{Eq. (22)}$$

where p denotes the dropout rate. This regularization method allows for random deactivation of neurons through training and aids in reducing co-adaption between neurons, thereby assisting with generalization performance.

3.3.6 SoftMax-Based Disease Prediction

Final probabilities are computed using the softmax function, as defined in equation (23):

$$P(y = i\,|x) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad \textit{Eq. (23)}$$

The predicted class label is determined using the maximum posterior probability, as expressed in equation (24):

$$\hat{y} = argmax_i P(y = i|x) \quad \textit{Eq. (24)}$$

The model confidence is therefore defined as the maximum softmax probability value associated with predicted loss.

3.3.7 Fine-Tuning Strategy

Selective fine-tuning can be used to improve accuracy for plant disease image classification tasks. To ensure stable training, we will unfreeze the last few layers of MobileNetV2 but keep all Batch Normalization layers frozen. We will apply a lower learning rate during this phase to preserve the feature representations learned from MobileNetV2 while minimizing the chance for overfitting.

Weights are updated via gradient descent, as defined in equation (25):

$$W_{t+1} = W_t - \eta \nabla L(W_t) \qquad Eq.\ (25)$$

Equation (26) expresses the employed categorical cross-entropy loss for multi-class classification.

$$L = -\sum_{i=1}^{C} \quad y_{iloglog\,(\widehat{y_i})} \qquad Eq.\ (26)$$

As formulated in equation (26), the loss penalizes the deviations among the projected probabilities and true class labels by controlling the optimization method.

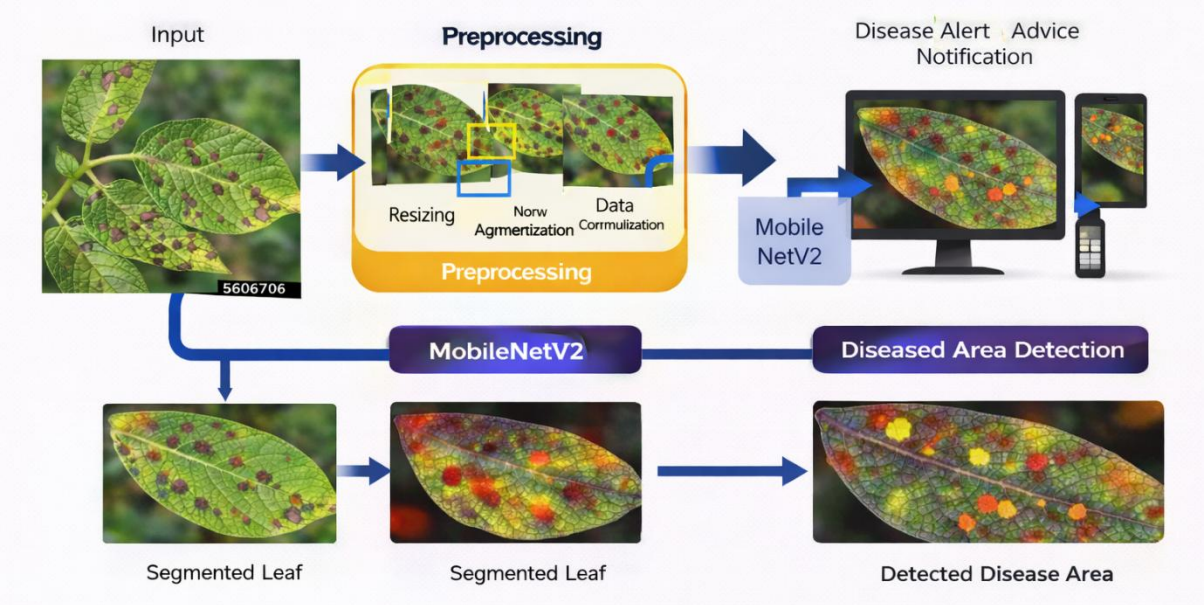


Figure 6: Structure of MobileNetV2 for Disease Detection.

The structural configuration of MobileNetV2 used in this study is illustrated in **Figure 6**. The architecture includes inverted residuals with linear bottlenecks, which improve feature propagation, but limit beginner loses. The architecture also allows for the network to learn

subtle visual features, such as lesion boundaries and colour changes, that are crucial to identifying the various attributes of a disease accurately.

The Adam optimization function work by using the 1st and 2nd instant of the gradient's estimates to create a model's parameters adjustment. The exponential moving average of the gradients is calculated in equation (27):

$$m_t = \beta_{1m_{t-1}} + (1 - \beta_1)g_t \quad \textit{Eq. (27)}$$

The determination of the squared gradients' moving averages is defined in equation (28):

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)g_t^2 \quad \textit{Eq. (28)}$$

The parameters are updated according to equation (29):

$$W_t = W_{t-1} - \frac{\eta m_t}{\sqrt{v_t} + \epsilon} \quad \textit{Eq. (29)}$$

3.4 Environmental Context Analysis

Depending on where a farmer lives, weather APIs are utilized to obtain different ecological issues such as temperature and moisture, which enable us to provide more reliable diagnosis and recommendations for farmers. These factors are processed along with the results of the disease prediction to form the suitability of the environment for the development of the disease. This approach integrates the environmental factors with image classification and helps improve disease prediction and eliminate false prediction. This is an additional factor that can be used to make suggestions for solutions and prevention strategies to protect plants.

Weather parameters are retrieved via APIs and modelled in equation (30) as:

$$E = \{T, H\} \quad \textit{Eq. (30)}$$

A disease risk score is computed in equation (31):

$$R = \alpha P_d + \beta T + \gamma H \quad \textit{Eq. (31)}$$

where P_d is disease probability. As indicated in equation (31), the risk score integrates both visual and environmental information, enabling a more reliable disease assessment.

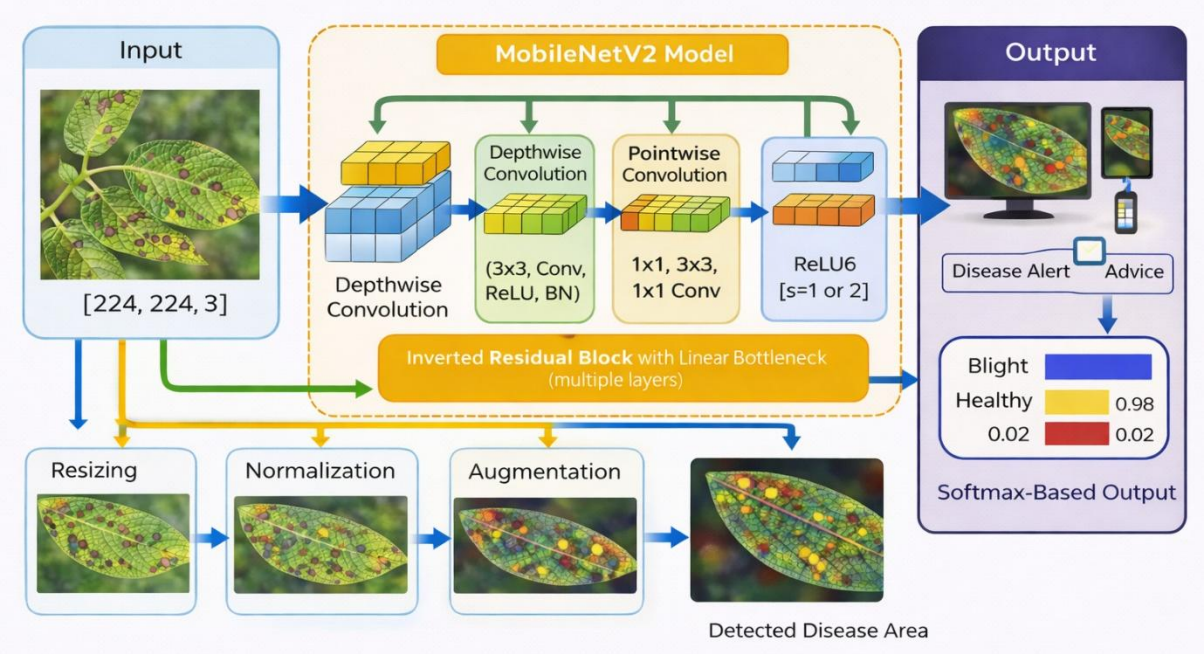


Figure 7: Proposed MobileNetV2 Architecture for Automated Potato Leaf Disease Detection.

A diagram showing how the integrated framework of MobileNetV2, environmental intelligence, and knowledge-based recommendations can create visual predictions and superimpose context data to provide actionable insights to convert traditional classifiers into intelligent decision-support systems can be seen in **Figure 7**.

3.5 Knowledge-Based Recommendation Generation

Disease-specific information is retrieved from an organized knowledge base related to agriculture after the disease has been identified and environment analysed. Information related to symptoms, causes, recommended organic and chemical treatments, dosage, and prevention strategies are included within the knowledge base. The system creates individual treatment and prevention strategies for diseases based on the disease and environment which prevails.

Recommendations are formulated as defined in equation (32):

$$Rec = f(Disease, R, K_b) \qquad Eq.\ (32)$$

where K_b is the knowledge base.

Similarity matching is expressed in equation (33) as:

$$Sim(q,k) = \frac{q \,.\, k}{\|q\| \, \|k\|} \qquad Eq.\ (33)$$

3.6 Alert Delivery and Farmer Notification

Diagnostic reports and recommendations will be sent to farmers and periodically updated. Farmers will receive e-mail notifications as well as alerts through the application interface, giving them timely updates regarding disease status along with advice. This gives them the opportunity to respond quickly by taking the necessary corrective measures to minimize losses and damage. Notification probability is defined in equation (34):

$$P_{alert} = \{1, \ R > \tau \quad 0, otherwise \qquad Eq. (34)$$

3.7 Feedback Collection and System Learning

Farmers are encouraged to continue to provide input on how effective the recommended treatments have been and how accurate the disease forecast has been to help us continue our research. Input is taken through the application interface and subsequently used to expand the knowledge base and improve our decision-making process. The learning stage based on this response allows for improved constancy, adaptableness, and correctness of predictions for the farmer's real-world agricultural conditions. The feedback updating mechanism of the system is expressed in equation (35):

$$K_b^{new} = K_b^{old} + \lambda F \qquad Eq. (35)$$

4. Experimental Setup

4.1 Implementation Environment

The Keras API on TensorFlow was used to create a plant disease identification model. To train the models efficiently, all experiments occurred on a local-machine and with a separate dedicated-GPU. Large datasets of images from the PlantVillage dataset were used for training, and the use of a GPU sped up the entire process.

We have developed a new MobileNetV2-based model for detecting diseases using nicely documented libraries provided by TensorFlow 2.x and Keras. The main advantages of TensorFlow included greater scalability, efficient use of GPUs, and robust support for multiple transfer learning approaches; consequently, we chose TensorFlow as our framework for developing this disease detection model.

4.2 Dataset Description

Training and Testing of the projected plant disease finding are utilized the PlantVillage dataset, which is a publicly available database that is commonly used as a benchmark dataset for agricultural image classification tasks. It contains a series of clear, high-quality images of

both healthy and disease damaged leaves from various types of crops (15 disease classes). All images were taken under controlled conditions to ensure uniform and recognizable patterns among the leaf images.

The PlantVillage dataset contains 20638 total images (15 different classes, i.e., diseased) allowing the model to learn different visual patterns associated with leaf diseases (for example, lesions, discoloration, blight, texture) across different crop types and varying degrees of severity of the disease.

To comply with the MobileNetV2 architecture's input size requirements, all images were resized to 224 x 224 pixels. Additionally, the pixel values were normalized using the Preprocess_input function to provide a consistent distribution of input pixel values compatible with the MobileNetV2 architecture during image classification after the transfer learning process begins, aided by using pretrained weights from ImageNet to aid with model convergence.

We created separate training and validation datasets by splitting the dataset as can be seen below: 4122 images in the validation dataset, and 16516 images in the training dataset, using an 80:20 split. The large number of images available for training enables the model to be trained more effectively using many types of data augmentation (rotations, zooms, width and height shifts, shear transforms, horizontal flips, etc.) while training to help prevent overfitting and increase generalization performance by creating many unique and different looking and sized training sample images.

Utilized the New PlantDisease Dataset in addition to the PlantVillage dataset to further investigate the efficiency of the proposed model. There are 38 classifications present within the New PlantDisease dataset; these classifications include healthy and diseased plant leaves.

In addition to the PlantVillage dataset, the FieldPlant dataset has 5170 images of leaf plants with 19 various leaf plant types. FieldPlant's dataset contains pictures taken in the field without being under any type of controlled circumstances, whereas both the PlantVillage and New PlantDisease datasets were created in controlled circumstances. Because the FieldPlant dataset contains the images of plant leaves under different factors that were not recorded, such as variable lighting, background clutter, leaf orientation, vegetable obstruction and background sounds, it will be more challenging to classify the leaves of plants from this dataset than those of the other two datasets. The symptoms of diseases seen in plants from this dataset are produced by real-life conditions in agricultural fields; thus, it will be more challenging to

classify leaves in the FieldPlant dataset and confirm the trustworthiness of our proposed model than in the other two datasets. Achieve the usefulness and applicability of the proposed method, for aggregate our three datasets together to measure the reliability and generalizability. Since the PlantVillage dataset consists of images of plant leaves taken under almost perfect controlled environments, it is critical to assess our model's performance within a practical context. Thus, the New PlantDisease and FieldPlant datasets provide variability within a real world, allowing us to better evaluate the overall performance of the proposed model while also having an equal number of plant disease types and balanced representation in the PlantVillage dataset, thus enabling creation of accurate deep learning classifiers for the identification of plant disease types.

4.3 Training Configuration

Training occurs in two phases:

i) Phase 1: Frozen backbone, trained classification head (10 epochs)

ii) Phase 2: Partial unfreezing with fine-tuning (10 epochs)

Key parameters:

i. Training Batch size: 32
ii. Initial learning rate: 0.001
iii. Fine-tuning rate: 0.00001
iv. Optimizer: Adam

Stability techniques include early stopping, learning rate decay, and checkpointing.

Training minimizes empirical risk which is defined in equation (36):

$$ERM = \frac{1}{N} \sum_{i=1}^{N} \quad L(y_i, f(x_i)) \qquad \textit{Eq. (36)}$$

The condition for early stopping is expressed in the equation (37):

$$L_{val}(t) > L_{val}(t-k) \qquad \textit{Eq. (37)}$$

The learning rate decay is formulated in the equation (38):

$$\eta_t = \eta_0 e^{-kt} \qquad \textit{Eq. (38)}$$

4.4 Model Complexity

Because of MobileNetV2 weighs substantially less than typical CNN models as well as having about 3.5-4 million learnt parameters and achieving similar overall performance statistics, it was selected for use in this study. The lightweight nature of the MobileNetV2 architecture allows it to be deployed on low resource devices and for use in agro-technology applications in real time.

5. Results and Performance Analysis

To evaluate the success of using an Agentic AI plant disease detection framework, a publicly available dataset consisting of labeled images of healthy and unhealthy plant leaves from multiple crop types was utilized. The dataset was split, with 80% being used for training the model and 20% as validation data, with the images being preprocessed prior to training the model. Preprocessing actions performed on the images included resizing, normalizing pixel intensity values, and augmenting the data to add variety within the features represented in the images. Preprocessing actions generated sufficient data to can evaluate how well the trained model will accurately generalize to new data.

5.1 Performance Evaluation on PlantVillage Dataset (Proposed Model)

The proposed framework leverages the MobileNetV2 network architecture and employs the PlantVillage dataset as a control for experimentation to accurately detect plant disease. The goal is to demonstrate accuracy in classifying plant types while minimizing the time it takes to compute results (maximize processing efficiency) which are appropriate for implementation in real time by farmers.

The MobileNetV2 model has reached an overall validation accuracy of 95.32%, which means it can generalize to many different crops and diseases with very little overfitting (crops and diseases, respectively). In fact, certain classifications, such as Potato Early Blight, Pepper Bell Healthy, and Tomato Yellow Leaf Curl Virus, scored nearly 100% for accuracy and demonstrated that the MobileNetV2 Model is capable of accurately identifying disease on crops regardless of their visual features. In addition, these classifications also exhibited a large amount of precision/recall and, therefore, large F1 values.

There were small drops in performance on visually similar classes (e.g., Tomato Early Blight vs. Target Spot), but this is offset by high overall average results (i.e., the average macro

and weighted scores are both close to 0.95), showing that overall learning was balanced and that there wasn't much class bias in terms of the precision/recall metrics calculated.

The results indicate that the MobileNetV2s is a lightweight and efficient architecture capable of accurately diagnosing plant disease while requiring less computational resources, making it ideal for agricultural applications.

5.1.1 Evaluation Metrics

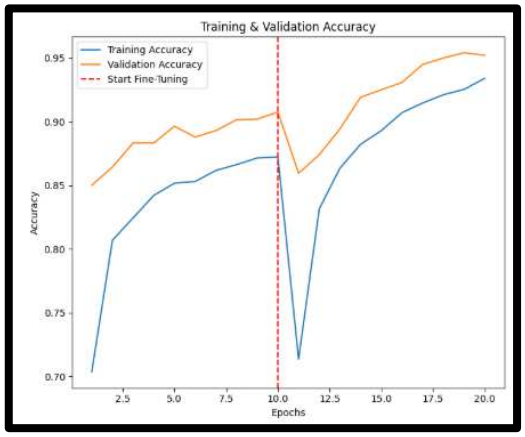


Figure 8: Evaluation Accuracy of Disease Detection Agent.

As depicted in **Figure 8**, the accuracy curves of MobileNetV2 trained on the PlantVillage dataset showed a notable increase in both training and validation accuracy at every epoch of training across all the classes. The proximity between training and validation accuracies throughout the training process indicates that the MobileNetV2 has a high degree of generalization for new or previously unseen data. The small amount of divergence between the training and validation accuracies also indicates that the MobileNetV2 did not overfit, indicating that the techniques of data augmentation, dropout regularization and fine-tuning were successfully applied. The smooth convergence pattern of the accuracy curves further demonstrates that the MobileNetV2 is effective for multi-class classification of plant disease and is due to stable updates of gradients during training and proper definition of the learning rate.

Commonly used classification metrics such as accuracy, precision, recall, F1 score and training loss were used to evaluate the model and provide insight into its predictive capability. Accuracy is defined in equation (39) as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad Eq.\ (39)$$

where

TP (True Positive): Accurate predictions of positive diseases

TN (True Negative): Accurate predictions of negative diseases

FP (False Positive): Inaccurate predictions of positive disease

FN (False Negative): Inaccurate predictions of negative disease

The model achieved an accuracy of 95.32%, demonstrating its effectiveness in learning discriminative disease features from leaf imagery.

The trained model recognized the distinction among diseased leaf images reasonably well, achieving an overall predictive accuracy of 95.32%. Evaluating the training loss and using that information to evaluate the optimization behaviour were necessary next steps in this analysis (as illustrated in **Fig. 9**). The overall trend of the loss curve was consistently decreasing and had only limited amounts of variation; thus, there were stable gradient updates and efficient minimization of prediction error. Furthermore, since training and validation loss never had a significant amount of divergence, we conclude that the model did not experience a great deal of overfitting while also having a good amount of generalizability.

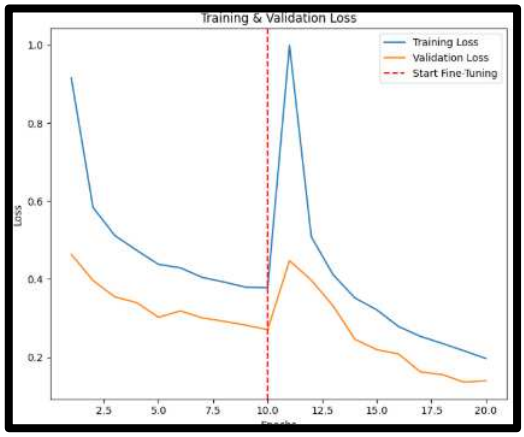


Figure 9: Training Loss of Disease Detection Agent.

In **Figure 9**, the training and validation loss curves of the proposed model are presented. The fact that the training loss curve was consistently trending downward indicates that prediction error was minimised effectively during the optimisation step. The validation loss curve shows a similar trend and does not show any extraordinary spikes or oscillations; therefore, we conclude that there was stable convergence performance. Some minor amounts of loss variation are consistent with the nature of deep learning training and does not indicate any lack of stability. The fact that there were no observable increases in the gap between the training and validation loss also validates that the model did not experience a large amount of

overfitting and confirms the two-phase training technique and learning rate schedule described in this work.

5.1.2 Class-wise Performance Analysis

A detailed classification report summarizing precision, recall, and F1-score for each disease category is presented in **Table 2**. Most classes achieved performance scores exceeding 0.95, underscoring the strong discriminative capacity of the MobileNetV2 architecture.

Notably, Tomato Yellow Leaf Curl Virus achieved the highest classification performance with an F1-score of 0.99, indicating near-perfect separability. Similarly, healthy plant categories exhibited high recall values, suggesting reliable differentiation between diseased and non-diseased samples.

Slightly lower performance was observed for Tomato Early Blight, which recorded an F1-score of 0.81. This reduction can be attributed to visual similarities with Tomato Late Blight, as both diseases exhibit overlapping symptoms such as necrotic lesions and concentric ring patterns. Such misclassifications are expected in fine-grained image recognition tasks where inter-class variance is minimal.

Table 2: Confusion Matrix of Proposed Model

Plant Classes	Precision	Recall	F1-Score	Support
Bacterial Spot (Bell Pepper)	0.97	0.96	0.97	199
Healthy (Bell Pepper)	0.98	0.98	0.98	295
Early Blight (Potato)	0.97	0.98	0.98	200
Late Blight (Potato)	0.96	0.97	0.97	200
Healthy (Potato)	0.97	0.93	0.95	30
Bacterial Spot (Tomato)	0.98	0.96	0.97	425
Early Blight (Tomato)	0.89	0.74	0.81	200
Late Blight (Tomato)	0.96	0.94	0.95	381
Leaf Mold (Tomato)	0.95	0.95	0.95	190
Septoria Leaf Spot (Tomato)	0.88	0.95	0.91	354
Spider Mites (Tomato)	0.92	0.96	0.94	335
Target Spot (Tomato)	0.94	0.87	0.9	280
Yellow Leaf Curl Virus (Tomato)	0.99	1	0.99	641
Mosaic Virus (Tomato)	0.89	0.97	0.93	74
Healthy (Tomato)	0.97	0.99	0.98	318

Table 2 demonstrates that the class level performance metrics consistently show high precision and recall for many types of diseased plants. Specifically, Tomato Yellow Leaf Curl

Virus had an F1 score very close to 1.0 providing confirmation that the features can be separated with a very high level of confidence.

Overall, the extremely high level of approximately 0.95 for the macro-average precision, recall and F1 scores across all 15 disease classes indicates that the overall predictive performance of the model is very consistently balanced. The model has been able to represent the many details of the different diseases present; for instance, the model was able to predict the presence of a disease by examining the characteristics of a lesion such as how the texture looks, what colour the lesions are, and what shape the lesions are.

In **Figure 10**, a confusion matrix was built to provide a class level view of the predicted results. The high count of values near the diagonal line indicates that most of the samples were classified correctly. The low count of values located off the diagonal line indicates that little confusion between the classes has occurred; however, in some instances (such as misclassifying Tomato Early Blight for Tomato Late Blight), visually similar categories produce considerable confusion between classes. This occurs because both categories show similar symptoms (necrotic spots and irregularly shaped lesion borders), thus reducing separability between categories.

In conclusion, the results of the confusion matrix provide support for the effectiveness of the proposed methodology and show similar performance across all categories of disease.

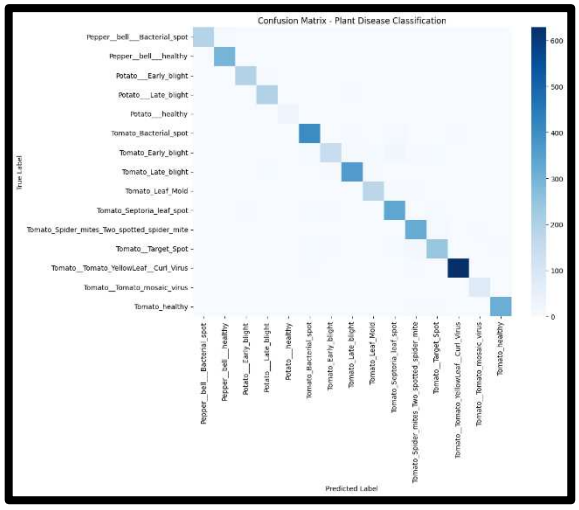


Figure 10: Confusion Matrix of Proposed Plant Disease Detection System.

Overall, the model demonstrated predictive balance among different types of diseases by achieving approximately the same macro-average precision, recall, and F1-score value of about

0.95 for an overall average predictive capability across multiple disease classifications. Table 3 represents overall performance metrics of the proposed model.

Table 3: Overall Performance Metrics of Proposed Model

Metric	Value
Accuracy	95.32%
Macro Precision	0.95
Macro Recall	0.95
Macro F1-score	0.95

5.1.3 Computational Efficiency and Deployment Feasibility

Moreover, computational efficiency was an important design goal along with prediction performance. MobileNetV2 uses depth wise separable convolutions which result in significantly fewer parameters relative to other CNNs.

This efficiency provides several practical advantages:

- Faster inference times
- Reduced memory consumption
- Lower power requirements
- Improved suitability for edge deployment

These characteristics position the proposed framework as a viable solution for real-time agricultural diagnostics, particularly in rural settings where high-performance computing infrastructure may not be readily available.

5.1.4 Practical Implications of the Agentic Framework

The system design has extended beyond traditional classification models by adding agent-based coordination into the overall technology ecosystem. As such, this system leverages disease prediction capabilities, evaluates risk factors from the environment, and provides knowledge-based solutions and recommendations to support proactive crop management.

This capability allows smallholder farmers to receive earlier intervention, thus reducing the likelihood of crop loss. Continuous feedback from actual use of the system enhances its adaptability, supporting ongoing, iterative improvements through machine learning.

5.1.5 Limitations and Future Scope

Although the PlantVillage dataset used to train and evaluate the model was derived from controlled conditions, real-world conditions may not mirror these same conditions due to a variety of factors (i.e., consistent light levels or backgrounds, objects passing in front of the camera during image capture, leaves partially covered by other leaves). All of these can contribute to biases associated with using the original dataset for training and evaluating models, and all of these may impact an overall system model's ability to generate accurate predictions.

For these reasons, future research efforts must focus on validating the system using actual agriculture databases and conducting controlled, large-scale agricultural field trials. Additional methods that should enhance the robustness of a model in real agricultural cultivation settings include the use of domain adaptation, multimodal sensing technologies, and real-time image-enhanced processes.

5.2 Comparative Analysis with ResNet50 and VGG16 (PlantVillage Dataset)

MobileNetV2 was compared with more complex convolutional networks, including ResNet50 and VGG16, to test the effectiveness of the original and proposed approaches. ResNet50 reached an accuracy of about 90.0% compared to MobileNetV2, showing that MobileNetV2 performed better than ResNet50. Although ResNet50 performed well on most classes of tomato diseases, the reduced recall rate on some specific tomato diseases led to a drop in the overall consistency of the model's performance. The higher complexity of ResNet50 also results in a large increase in the computational resources needed for optimal performance compared to MobileNetV2.

The confusion matrix shows how well the ResNet50 model performs with classifying images in PlantVillage, as seen in **Figure 11**. The rows and columns of the confusion matrix show the actual and predicted classes (in this case, the classes of plant diseases). Thus, the confusion matrix provides a summary of how accurately the classification model performed. The elements on the diagonal of the matrix are the number of instances that were classified correctly, while the elements not on the diagonal are the number of instances classified incorrectly. Thus, the confusion matrix clearly shows that the ResNet50 model can accurately identify instances of each plant disease class, as evidenced by the significant number of elements located along the diagonal of the matrix.

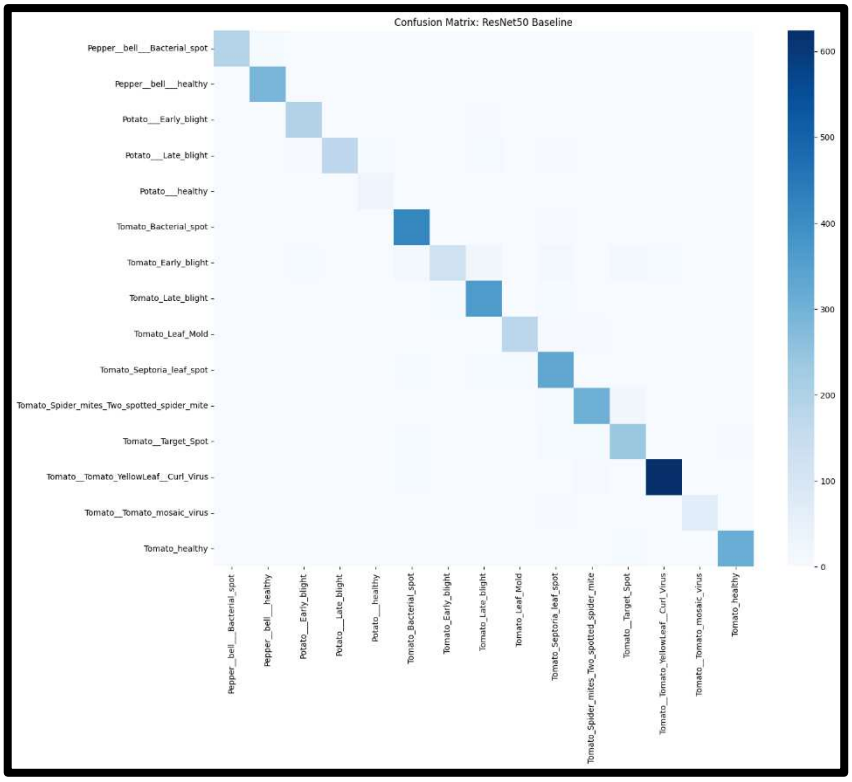


Figure 11: Confusion Matrix of ResNet50 Model.

The ResNet50 model has a high level of accuracy in the prediction of most classes, especially those plant diseases that have unique visual appearances. The low level of confusion between classes indicates that the model can effectively capture the discriminative features of the leaf images of plants. However, some misclassifications are observed between classes that have similar visual appearances, which can be attributed to the overlapping texture patterns and colour differences.

Table 4: Class-wise Performance Metrics of ResNet50 Algorithm in PlantVillage Dataset

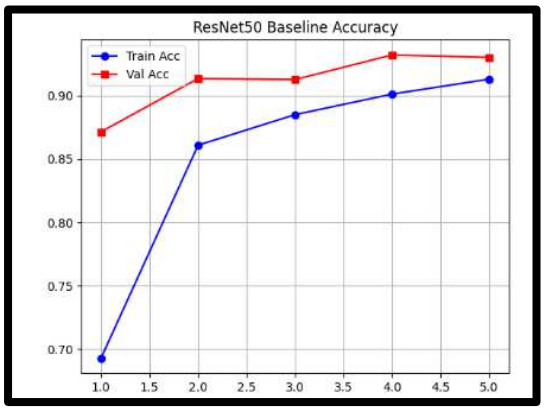
Plant Classes	Precision	Recall	F1-Score	Support
Bacterial Spot (Bell Pepper)	0.9641	0.9447	0.9543	199
Healthy (Bell Pepper)	0.9699	0.9831	0.9764	295
Early Blight (Potato)	0.9372	0.97	0.9533	200
Late Blight (Potato)	0.9884	0.85	0.914	200
Healthy (Potato)	0.7941	0.9	0.8438	30
Bacterial Spot (Tomato)	0.9205	0.9812	0.9499	425
Early Blight (Tomato)	0.9197	0.63	0.7478	200
Late Blight (Tomato)	0.8856	0.9554	0.9192	381
Leaf Mold (Tomato)	0.9365	0.9316	0.934	190
Septoria Leaf Spot (Tomato)	0.8571	0.9322	0.8931	354
Spider Mites (Tomato)	0.9159	0.9104	0.9132	335
Target Spot (Tomato)	0.8427	0.8607	0.8516	280
Yellow Leaf Curl Virus (Tomato)	0.9889	0.9735	0.9811	641

Mosaic Virus (Tomato)	0.9452	0.9324	0.9388	74
Healthy (Tomato)	0.9842	0.9811	0.9827	318

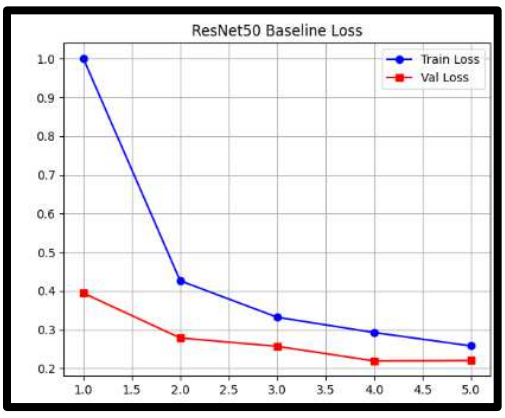
Table 4 provides class wise analysis of ResNet50 including precision, recall and f1 scores. Precision measures the amount of correct positive predictions, while recall measures how well the model detects all relevant instances in each class. The F1 score combines precision with recall providing a more balanced view on classification performance overall.

ResNet50 performs well across most of the disease classes, demonstrating it is generalizing well on complex agricultural images. Metrics for classes with higher support value are more stable and demonstrate that having enough training sample improves predictive accuracy. The lower numbers for a few classes may be attributed to class imbalance or visual similarity between diseases which are both long-standing issues in plant pathology datasets. Each of these class level metrics support the overall conclusion that ResNet50 is a good performing, deep learning model that can be used to accurately classify plant diseases.

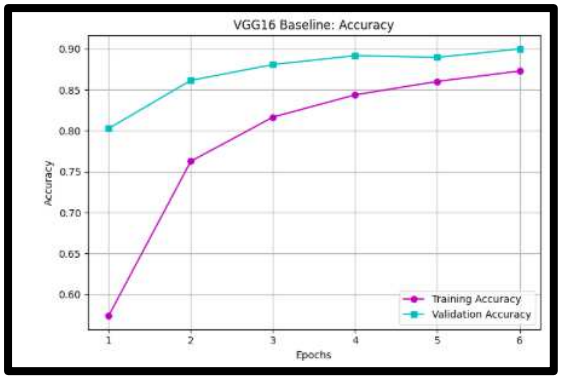
The training dynamics of the baseline ResNet50 and VGG16 models are shown in **Figure 12** through the accuracy and loss curves over several epochs. The progression of both the training and validation accuracy (shown in subfigures a and c) confirms the ability of both models to learn discriminative features from the plant disease data. ResNet50 has a quicker rate of convergence with a consistent increase in validation accuracy, which suggests its good generalization capability. The VGG16 also has a gradual increase in accuracy; however, its learning curve stabilizes later than that of ResNet50.

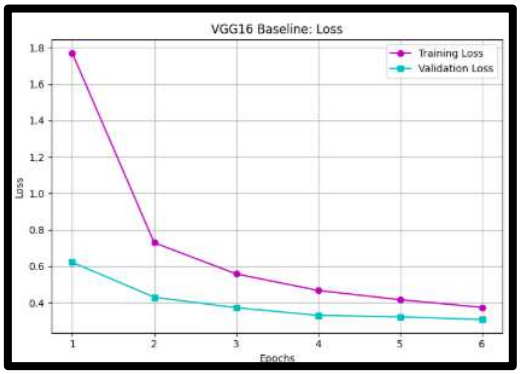


(a)



(b)





(c) **(d)**

Figure 12: Training, Validation Accuracy and Loss (a) Baseline Accuracy of ResNet50 Model, (b) Baseline Loss of ResNet50 Model, (c) Baseline Accuracy of VGG16 Model, (d) Baseline Loss of VGG16 Model.

The loss curves are shown in subfigures b and d. The decline through the validation loss curves provided evidence of the predicted errors being minimized by both models. The small distance between the training and validation losses suggests little overfitting, indicating that the training process was robust. Overall, the results indicate both architectures will perform well when classifying plant diseases; however, ResNet50 was better optimally trained and stable.

Based on comparative analysis of these methods, it has been demonstrated that the less complex architecture achieved overall superior accuracy in tasks related to the identification of disease types associated with agricultural plants than did either of the more involved neural networks studied (ResNet50 and VGG16). The MobileNetV2 was shown to perform close to ResNet50 for correctly classifying samples (i.e., predicting diseased leaves) but requires additional calculation capacity to process an increased number of model parameters relative to its more complex counterpart.

The confusion matrix related to the VGG16 model (**Figure 13**) gives a summary of performance across multiple classes of plant disease from the PlantVillage dataset by displaying the model's true positive (TP) and false positive (FP) classifications at the summary level. A large proportion of the total sample classifications fall on the major diagonal within this matrix. This indicates that the VGG16 architecture has successfully differentiated between plant leaves with visible (i.e., healthy) symptoms versus those with invisible (i.e., diseased) solutions to an extent greater than chance would allow.

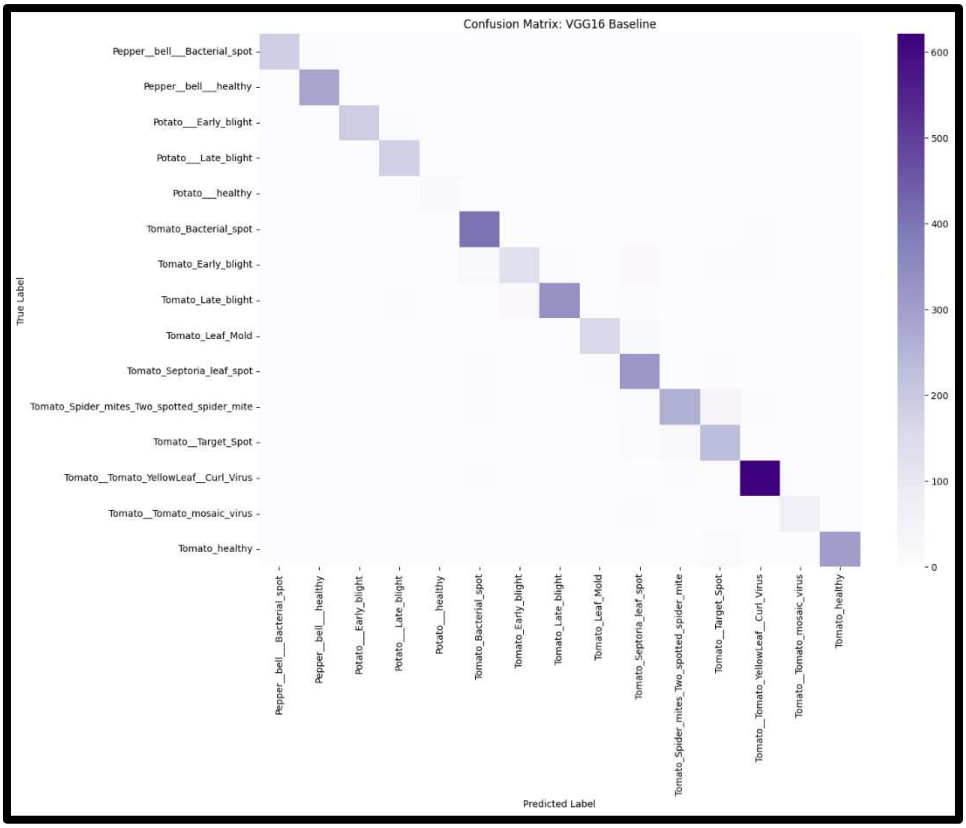


Figure 13: Confusion Matrix of VGG16 Model.

The performance of the model shows great strength for predicting the classes with Tomato Yellow Leaf Curl Virus (TYLCV), Pepper bell healthy (PBH) and Potato Early blight (PEB) since the number of correctly classified instances significantly exceeds those that were misclassified. However, there are several diseases that display some degree of confusion, mainly due to close visual similarity such as Tomato Early blight and Target Spot having moderate confusion when identified through the trained model. The results of the confusion matrix confirm that VGG16 continues to demonstrate consistent multi-class classification performance over different classes as well as sufficient ability to interpret different plant leaf images based on complex representations of features.

As shown in **Table 5**, using class-wised evaluation metrics such as precision, recall, F1 score and support allows for a more thorough understanding of the predictive ability of the algorithm being evaluated. In creating the predictive model, high precision values were found within classes including Tomato Healthy (.9744), Pepper bell healthy (.9568) and Tomato Yellow Leaf Curl Virus (0.9569), meaning these classes had very few false positives created by the algorithm. The model also returned a high number of recall scores across most classes, suggesting that the model was able to correctly determine the presence of disease on plant leaves contained within the training data.

However, comparatively lower recall in classes such as Potato healthy (0.7333) and Tomato Early blight (0.655) suggests that some true samples were misclassified, potentially due to limited training data or visual resemblance between disease categories. The F1-scores across most classes remain consistently high, reflecting a balanced trade-off between precision and recall. Additionally, classes with larger support values, particularly Tomato Yellow Leaf Curl Virus, show stable performance, emphasizing the importance of sufficient data for deep learning models.

VGG16's class-specific metrics indicate that VGG16 is a solid convolutional neural network and can accurately classify plant diseases, but future improvements in performance could occur due to enhanced balancing and augmentation of the datasets used for training.

Table 5: Class-wise Performance Metrics of VGG16 Algorithm in PlantVillage Dataset

Plant Classes	Precision	Recall	F1-Score	Support
Bacterial Spot (Bell Pepper)	0.931	0.9497	0.9403	199
Healthy (Bell Pepper)	0.9568	0.9763	0.9664	295
Early Blight (Potato)	0.9598	0.955	0.9574	200
Late Blight (Potato)	0.9091	0.9	0.9045	200
Healthy (Potato)	0.9565	0.7333	0.8302	30
Bacterial Spot (Tomato)	0.8945	0.9576	0.925	425
Early Blight (Tomato)	0.8037	0.655	0.7218	200
Late Blight (Tomato)	0.9382	0.8766	0.9064	381
Leaf Mold (Tomato)	0.9138	0.8368	0.8736	190
Septoria Leaf Spot (Tomato)	0.801	0.9096	0.8519	354
Spider Mites (Tomato)	0.8829	0.7881	0.8328	335
Target Spot (Tomato)	0.7524	0.8357	0.7919	280
Yellow Leaf Curl Virus (Tomato)	0.9569	0.9688	0.9628	641
Mosaic Virus (Tomato)	0.8289	0.8514	0.84	74
Healthy (Tomato)	0.9744	0.9591	0.9667	318

The MobileNetV2 architecture has the best trade-off between predictive accuracy and computational efficiency due to the use of depth wise separable convolutions in the architecture, thereby decreasing the total number of model parameters while preserving the fine details, which will make this model highly efficient for use in real-time agricultural applications.

5.3 Cross-Dataset Generalization Study

MobileNetV2 has shown that a lightweight architecture can achieve superior accuracy in plant disease classification when optimized, even when compared to larger networks.

Classifying plant diseases using MobileNetV2 demonstrated that a lightweight architecture can be highly accurate if it is optimized correctly. The results of generalization capability were determined by using the MobileNetV2 to perform cross-dataset evaluation across multiple datasets.

To validate that MobileNetV2 is an effective solution and can be used on more than just one dataset, additional experiments using a variety of plant disease datasets were conducted. Through these experiments, comparisons between performance were performed across varying distributions of data rather than being used as replacements for the original proposed experiments. The performance of a trained MobileNetV2 was further evaluated on the New PlantDisease Dataset for generalization on various datasets using confusion matrix and using individual class metrics.

Overall, the results provide evidence that the proposed architecture does not have a fixed performance profile on controlled datasets and that it demonstrates a robust capability for producing reliable performance when used in an agricultural environment.

5.3.1 Results on New PlantDisease Dataset

The purpose behind demonstrating MobileNetV2's strength as a broad framework and performance beyond one data set was tested through other plant disease data sets using additional experimentation. The experiments would have a focus on comparing efficiencies of different distributions of the same dataset, not as replacement for original proposed configurations. To assess the ability of the model to generalize on multiple distinct datasets, MobileNetV2 was trained and tested using the New Plant Disease dataset, and results were derived based on confusion matrix analysis and detailed class-wise metrics.

MobileNetV2 achieved excellent overall validation accuracy (97.05%) within the context of all samples used in the analysis; this indicates that MobileNetV2 is scalable across several crop types. In addition, all but a few of the classes achieved precision and recall greater than (0.95); thus, the results suggest that MobileNetV2 has been capable of creating stable feature extraction processes on heterogeneous condition for all identified plant diseases.

These and other results document that MobileNetV2 provides improved performance in terms of generalized performance and reduced overfitting compared to other methods (i.e., successful generalization) as MobileNetV2 has provided greater macro and weighted averages

than other deep learning methods (e.g. overfitting) across all datasets evaluated. **Table 6** presents overall performance metrics of Algorithms in New PlantDisease Dataset.

Table 6: Overall Performance Metrics of Algorithms in New PlantDisease Dataset

Model	Validation Accuracy
MobileNetV2	97.05%
ResNet50	95.66%
VGG16	93.44%

The performance of the MobileNetV2 model by class is illustrated in **Table 7** using metrics of precision, recall, F1-score, and support. This information provides an overall assessment of the model's ability to predict. Precision and recall are both very high for many of the classes, including Apple healthy (0.994), Corn healthy (0.997), Strawberry leaf scorch (0.998) and orange citrus greening (0.999); thus, indicating that the model had low rates of both false positive and false negative predictions.

Most of the classes have an F1-score greater than 0.97, indicating the relationship between precision and recall is well-balanced in the predictions made using the model. Another observation is that the model performed very well with diseases of high support; therefore, having adequate training examples greatly improved the stability of the classification made with the model. For example, classes like Potato Early blight and Pepper bacterial spot demonstrate almost complete detection by the model; thus, indicating the strength of the model.

Table 7: Class-wise Performance Metrics of MobileNetV2 Algorithm in New PlantDisease Dataset

Plant Classes	Precision	Recall	F1_Score	Support
Scab (Apple)	0.970646	0.984127	0.97734	504
Black Rot (Apple)	0.986111	1	0.993007	497
Cedar Apple Rust (Apple)	0.993197	0.995455	0.994325	440
Healthy (Apple)	0.994024	0.994024	0.994024	502
Healthy (Blue Berry)	0.982609	0.995595	0.989059	454
Powdery Mildew (Cherry)	0.99284	0.988124	0.990476	421
Healthy (Cherry)	1	0.991228	0.995595	456
Cercospora Leaf Spot (Corn)	0.782524	0.982927	0.871351	410
Common Rust (Corn)	0.99375	1	0.996865	477
Northern Leaf Blight (Corn)	0.983784	0.763103	0.859504	477
Healthy (Corn)	0.995717	1	0.997854	465
Black Rot (Grape)	0.993492	0.970339	0.981779	472
Black Measles (Grape)	0.971487	0.99375	0.982492	480
Leaf Blight (Grape)	1	0.995349	0.997669	430
Healthy (Grape)	0.997636	0.997636	0.997636	423
Citrus Greening (Orange)	1	0.998012	0.999005	503

Bacterial Spot (Peach)	0.993407	0.984749	0.989059	459
Healthy (Peach)	0.979592	1	0.989691	432
Bacterial Spot (Bell Pepper)	0.979508	1	0.989648	478
Healthy (Bell Pepper)	1	0.971831	0.985714	497
Early Blight (Potato)	1	0.983505	0.991684	485
Late Blight (Potato)	0.991453	0.956701	0.973767	485
Healthy (Potato)	0.986577	0.967105	0.976744	456
Healthy (Raspberry)	1	0.98427	0.992072	445
Healthy (Soy Bean)	0.978558	0.994059	0.986248	505
Powdery Mildew (Squash)	0.993135	1	0.996556	434
Leaf Scorch (Strawberry)	1	0.997748	0.998873	444
Healthy (Strawberry)	0.991266	0.995614	0.993435	456
Bacterial Spot (Tomato)	0.956422	0.981176	0.968641	425
Early Blight (Tomato)	0.931507	0.85	0.888889	480
Late Blight (Tomato)	0.952273	0.904968	0.928018	463
Leaf Mold (Tomato)	0.90495	0.97234	0.937436	470
Septoria Leaf Spot (Tomato)	0.91762	0.919725	0.918671	436
Spider Mites (Tomato)	0.911111	0.942529	0.926554	435
Target Spot (Tomato)	0.867647	0.90372	0.885316	457
Yellow Leaf Curl Virus (Tomato)	0.997877	0.959184	0.978148	490
Mosaic Virus (Tomato)	0.988688	0.975446	0.982022	448
Healthy (Tomato)	0.950298	0.993763	0.971545	481

While the model has demonstrated good overall performance, Corn Northern Leaf Blight (0.763) and Tomato Early blight (0.850) had lower recall rates; thus, indicating that the model misclassified some samples. Some of this misclassification is likely due to characteristics of the samples. Specifically, samples may exhibit similarities, backgrounds conditions may be exceptionally complex, or disease symptoms may show slight differences; therefore, feature extraction may be challenging.

In summary, the use of MobileNetV2 demonstrates an excellent trade-off between accuracy and computational cost. In addition to the lightweight architecture of the model, the model made accurate predictions with high reliability, thus providing excellent potential for use in crop monitoring systems that are mobile-based and precision agriculture environments.

Moreover, examining the confusion matrix reinforces the reliability of the MobileNetV2 model by demonstrating a clear support for many disease classes. Close inspection of the confusion matrix shows clear support for many of the predicted classes along the diagonal of the matrix indicating that most of all predictions are correctly assigned to their respective classes. Misclassification occurrences are primarily between visually similar diseases specifically among leaf blight and spot diseases, where there are many disease features are

similar for example necrotic lesions, chlorosis and texture variation but are offset by different disease features. While there are some challenges with accurately classifying visually similar diseases using the MobileNetV2 model, the low off-diagonal error distribution provides confirmation of the model's excellent discriminatory capability! Due to the design of the MobileNetV2 architecture however, the model can perform efficient feature extraction via a lightweight depth wise separable convolutional structure while maintaining fine-scale spatial detail required for strong consistency between classes. Overall, these findings demonstrate that the proposed method is, therefore, both computationally efficient and very reliable for multi-class plant disease classification in a practical agricultural setting.

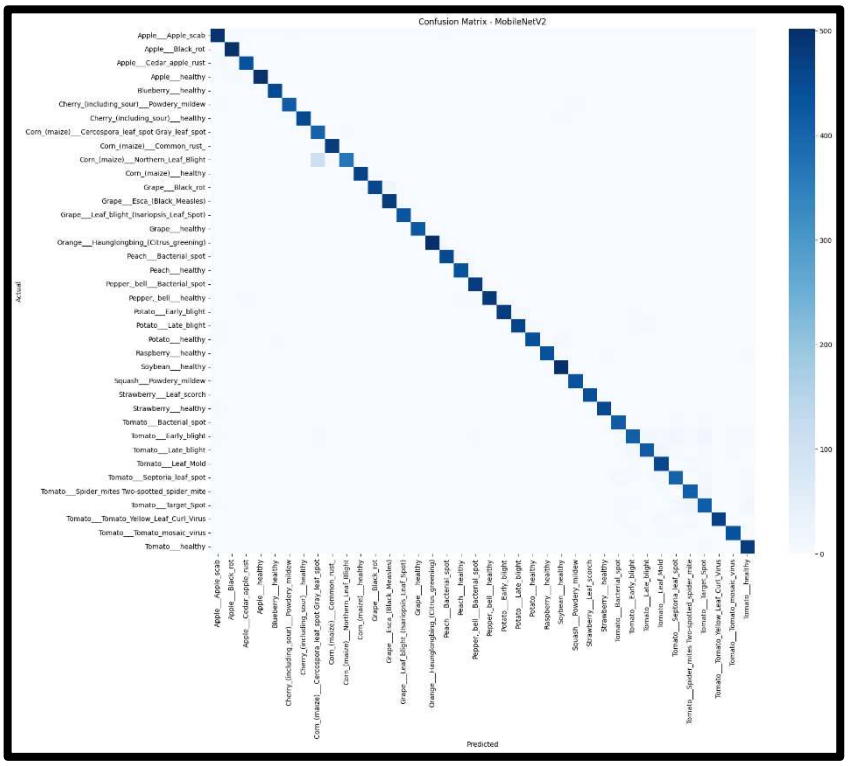


Figure 14: Confusion Matrix of MobileNetV2 Model of New PlantDisease Dataset.

The MobileNetV2 model's confusion matrix can be viewed in **Figure 14**, which shows how well it classified the various plant disease classes in the New PlantDisease dataset. There are a lot of numbers on the confusion matrix's diagonal, indicating that most of the samples were accurately classified. This confirms that MobileNetV2 is successful at extracting discriminative features while also being computationally efficient.

The model was able to recognize every class very well, except for a few. It had minimal misclassification between Apple Healthy, Orange Citrus Greening, Strawberry Leaf Scorch and Corn Common Rust. The high predominance on the diagonal of the confusion matrix also

indicates that although MobileNetV2 has fewer parameters than much larger architectures, it is still capable of producing high quality classification results.

Table 8: Class-wise Performance Metrics of ResNet50 Algorithm in New PlantDisease Dataset

Plant Classes	Precision	Recall	F1_Score	Support
Scab (Apple)	0.9427	0.9802	0.9611	504
Black Rot (Apple)	0.9876	0.9618	0.9745	497
Cedar Apple Rust (Apple)	0.9954	0.975	0.9851	440
Healthy (Apple)	0.9646	0.9761	0.9703	502
Healthy (Blue Berry)	0.9496	0.9956	0.972	454
Powdery Mildew (Cherry)	0.9906	0.9976	0.9941	421
Healthy (Cherry)	0.9956	0.989	0.9923	456
Cercospora Leaf Spot (Corn)	0.8879	0.9463	0.9162	410
Common Rust (Corn)	0.9979	0.9937	0.9958	477
Northern Leaf Blight (Corn)	0.9511	0.8973	0.9234	477
Healthy (Corn)	0.9979	1	0.9989	465
Black Rot (Grape)	0.9888	0.9386	0.963	472
Black Measles (Grape)	0.9444	0.9917	0.9675	480
Leaf Blight (Grape)	1	0.986	0.993	430
Healthy (Grape)	0.9953	0.9976	0.9965	423
Citrus Greening (Orange)	1	0.996	0.998	503
Bacterial Spot (Peach)	0.9719	0.9782	0.975	459
Healthy (Peach)	0.9882	0.9653	0.9766	432
Bacterial Spot (Bell Pepper)	0.946	0.9895	0.9673	478
Healthy (Bell Pepper)	0.9814	0.9577	0.9695	497
Early Blight (Potato)	0.9774	0.9814	0.9794	485
Late Blight (Potato)	0.9687	0.9567	0.9627	485
Healthy (Potato)	0.9579	0.9474	0.9526	456
Healthy (Raspberry)	0.9865	0.9888	0.9877	445
Healthy (Soy Bean)	0.9647	0.9743	0.9695	505
Powdery Mildew (Squash)	0.9977	1	0.9988	434
Leaf Scorch (Strawberry)	0.9889	1	0.9944	444
Healthy (Strawberry)	0.9978	0.9781	0.9878	456
Bacterial Spot (Tomato)	0.9344	0.9718	0.9527	425
Early Blight (Tomato)	0.8674	0.8313	0.8489	480
Late Blight (Tomato)	0.8972	0.905	0.9011	463
Leaf Mold (Tomato)	0.9293	0.9234	0.9264	470
Septoria Leaf Spot (Tomato)	0.9337	0.8073	0.8659	436
Spider Mites (Tomato)	0.9564	0.7563	0.8447	435
Target Spot (Tomato)	0.7585	0.8796	0.8146	457
Yellow Leaf Curl Virus (Tomato)	0.9583	0.9857	0.9718	490
Mosaic Virus (Tomato)	0.9387	0.9576	0.9481	448
Healthy (Tomato)	0.9042	0.9813	0.9412	481

There was, however, a fair amount of confusion between the visually similar class pairings of Corn Northern Leaf Blight vs. Cercospora Leaf Spot and Tomato Early Blight vs. Target Spot. The confusion between these class pairings is likely due to the overlap of symptoms that are present in each class (e.g., the shape of the lesions and the colour disruptions and patterns are all very similar). Therefore, it is evident from the MobileNetV2 confusion matrix that it provides reliable multi-class classification results and is thus an excellent candidate for real-time usage in agriculture where both speed and accuracy are essential.

To further substantiate the robustness of the MobileNetV2 model for classifying the New PlantDisease dataset, ResNet50's classification performance was assessed with both a confusion matrix and by calculating class-specific classification metrics

As presented in **Table 8**, the model achieves very high precision, recall, f1-score, and support values across most classes of plants examined; with some classes surpassing a f1-score value of 0.97 suggesting little to no imbalance between false-positive and false-negative classifications. For example, the model scored exceptionally well with Corn common rust (0.9958), healthy grape (0.9965), orange greening (0.998), and Powdery Mildew Squash (.9988), demonstrating excellent generalization capabilities by the model. The Tomato Spider Mite and Tomato Early Blight classes reported relatively lower recall values (.7563 and .8313 respectively). However, all the metrics remain very competitive while indicating stable classification performance, even in very complicated visual classifications (low viewability). The performance results of this study further confirm that ResNet50 is effectively capable of providing high levels of accuracy and dependability as an automated detection system for plant diseases (at large scale) and to provide for precision agriculture processes.

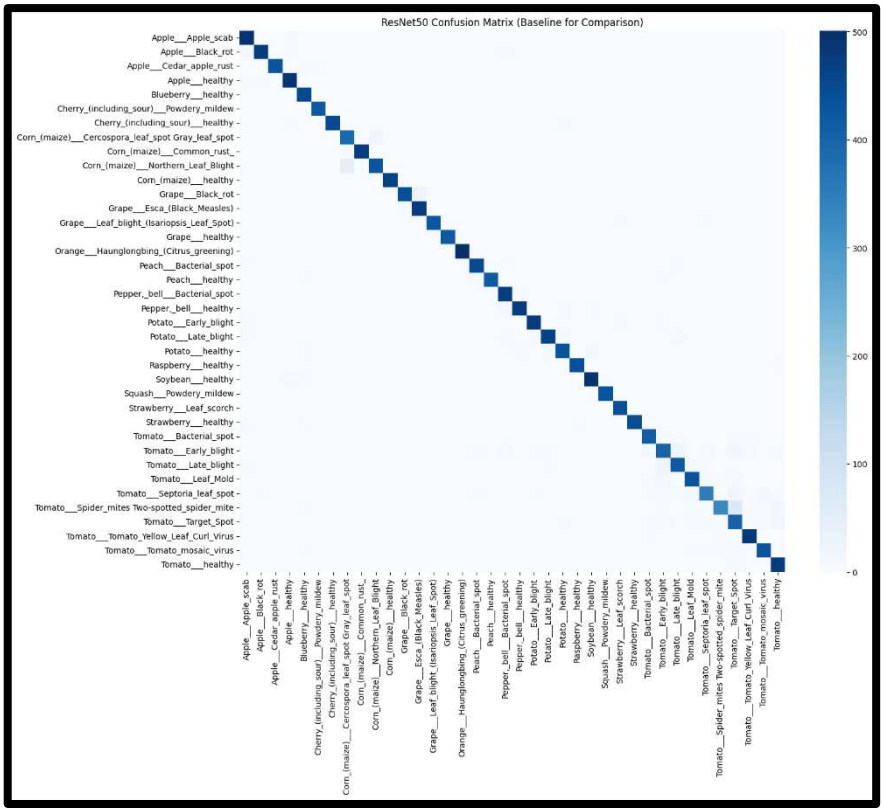


Figure 15: Confusion Matrix of ResNet50 Model of New PlantDisease Dataset.

The ResNet50 model's performance on the New Plant Disease Dataset can be seen in the confusion matrix shown in **Figure 15**. This figure visualises how well ResNet50 classified each plant disease category. Most classifications are in the diagonal and were correctly classified. Therefore, the model was able to learn complex visual patterns associated with each plant disease category and exhibit good prediction accuracy to many of the samples in the Corn healthy class, orange citrus greening class, Squash powdery mildew class, and Strawberry leaf scorch class, where instances of misclassification are almost non-existent. Thus, the ability of ResNet50 to extract deep features from images describes its capability to compare minor differences in leaf texture, colour, and structural characteristics of lesions. Although there was some confusion amongst the multiple categories of Tomato diseases (i.e., Early blight, Septoria leaf spot and Spider mites) due to overlapping visual symptoms and infection characteristics, the confusion matrix indicates that ResNet50 has produced very good multi-class classification results based on an unprecedentedly diverse and highly challenging dataset.

To further investigate task of classifying images of plants using the New Plant Disease Dataset, confusion matrices and class-specific evaluation metrics were used.

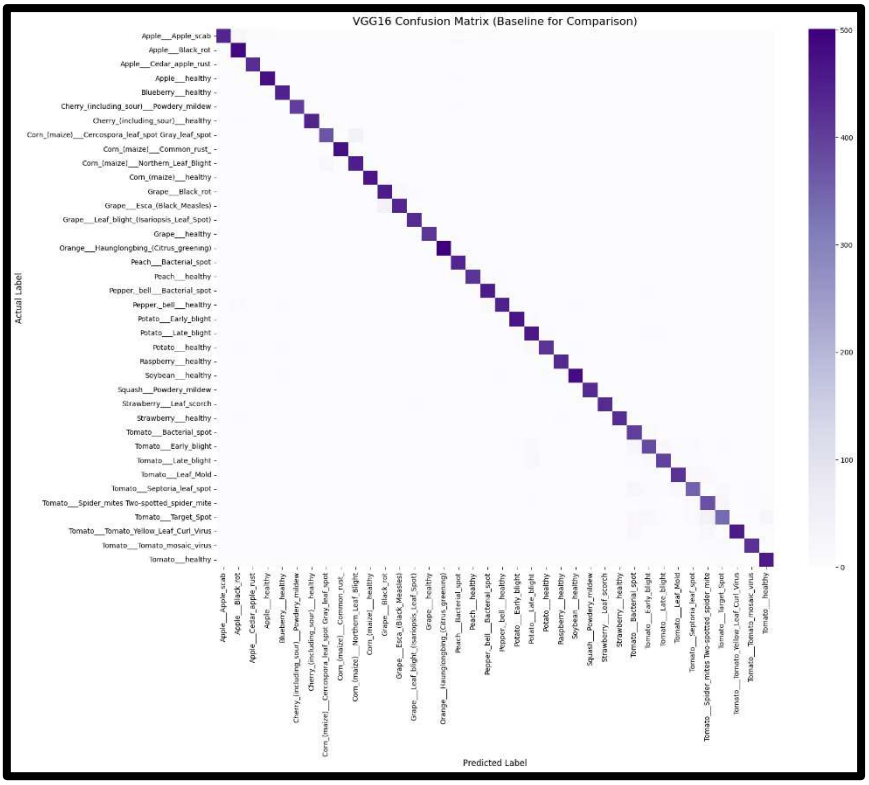


Figure 16: Confusion Matrix of VGG16 Model of New PlantDisease Dataset.

Figure 16 includes a grid of confusion matrices generated by applying VGG16 to the data set. This figure indicates that VGG16 classifies plant disease images accurately across many different types of plants. The numbers on the diagonal of this matrix are uniformly distributed and indicate that a majority are well-classified cases, which suggests that VGG16 does an adequate job of extracting hierarchical visual features of leaf images. In several specific categories (e.g., Corn common rust, Corn healthy, Squash powdery mildew, orange citrus greening), the distributions are very tightly clustered around the diagonal, indicating a very strong ability of VGG16 to produce highly consistent classifications based on the separability of features associated with the category.

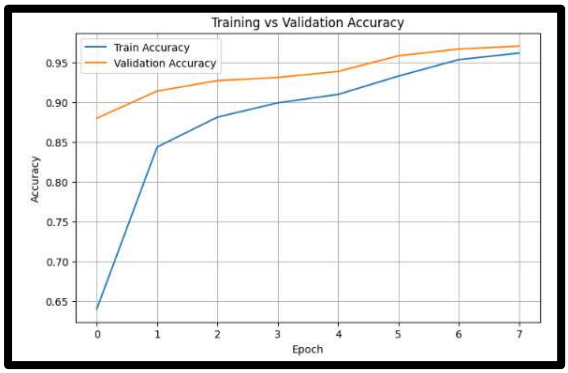
On the other hand, there are also instances of vague classifications within a group of three features represented by the classes of tomato disease (i.e., Early blight, Target spot, Septoria leaf spot), which is likely due to the similarity in the shapes and colours of the lesions caused by each class of disease, coupled with the timing of the progression of each class of disease. Overall, the results presented in the confusion matrix demonstrate that VGG16 can provide consistent multi-class recognition capacity and will recognize both healthy and diseased plants reliably.

Table 9: Class-wise Performance Metrics of VGG16 Algorithm in New PlantDisease Dataset

Plant Classes	Precision	Recall	F1-Score	Support
Scab (Apple)	0.9563	0.869	0.9106	504
Black Rot (Apple)	0.8743	0.9799	0.9241	497
Cedar Apple Rust (Apple)	0.9749	0.9705	0.9727	440
Healthy (Apple)	0.9557	0.9462	0.951	502
Healthy (Blue Berry)	0.9008	0.9802	0.9388	454
Powdery Mildew (Cherry)	0.9975	0.9501	0.9732	421
Healthy (Cherry)	0.9889	0.9737	0.9812	456
Cercospora Leaf Spot (Corn)	0.946	0.8976	0.9212	410
Common Rust (Corn)	0.9979	0.9979	0.9979	477
Northern Leaf Blight (Corn)	0.913	0.9455	0.9289	477
Healthy (Corn)	0.9915	1	0.9957	465
Black Rot (Grape)	0.9042	0.9597	0.9311	472
Black Measles (Grape)	0.9712	0.9125	0.9409	480
Leaf Blight (Grape)	0.9576	0.9977	0.9772	430
Healthy (Grape)	0.988	0.974	0.981	423
Citrus Greening (Orange)	0.9921	0.996	0.994	503
Bacterial Spot (Peach)	0.9224	0.9586	0.9402	459
Healthy (Peach)	0.9654	0.9676	0.9665	432
Bacterial Spot (Bell Pepper)	0.9401	0.9519	0.9459	478
Healthy (Bell Pepper)	0.945	0.8994	0.9216	497
Early Blight (Potato)	0.9529	0.9588	0.9558	485
Late Blight (Potato)	0.8598	0.9485	0.902	485
Healthy (Potato)	0.963	0.9145	0.9381	456
Healthy (Raspberry)	0.9709	0.9753	0.9731	445
Healthy (Soy Bean)	0.9505	0.9505	0.9505	505
Powdery Mildew (Squash)	0.9954	0.9954	0.9954	434
Leaf Scorch (Strawberry)	0.9954	0.9662	0.9806	444
Healthy (Strawberry)	0.9817	0.943	0.962	456
Bacterial Spot (Tomato)	0.8204	0.9459	0.8787	425
Early Blight (Tomato)	0.8297	0.8021	0.8157	480
Late Blight (Tomato)	0.8946	0.8618	0.8779	463
Leaf Mold (Tomato)	0.9211	0.8936	0.9071	470
Septoria Leaf Spot (Tomato)	0.8314	0.8028	0.8168	436
Spider Mites (Tomato)	0.8134	0.8621	0.8371	435
Target Spot (Tomato)	0.8204	0.7396	0.7779	457
Yellow Leaf Curl Virus (Tomato)	0.9806	0.9286	0.9539	490
Mosaic Virus (Tomato)	0.9548	0.942	0.9483	448
Healthy (Tomato)	0.927	0.9501	0.9384	481

The comparison of three deep learning models in training and validation accuracy and loss using New PlantDisease dataset is displayed in **Figure 17**. MobileNetV2 has a gradual upward trend for the validation accuracy and only experiences small fluctuations which indicates

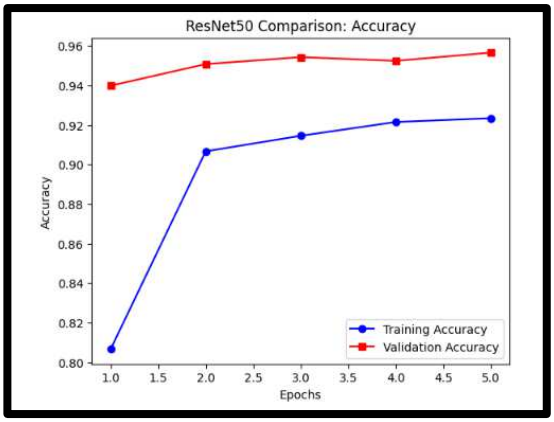
effective feature extraction and very low levels of overfitting between the two; therefore, there is clearly an improvement in the MobileNetV2.By comparison, the ResNet50 architecture is better defined as having the highest level of stability among the three architectures since training accuracy greatly improved during the beginning epochs and the continued downward trend in the loss graph for ResNet50 demonstrated the fastest convergence rate and the greatest optimization level throughout all epochs.



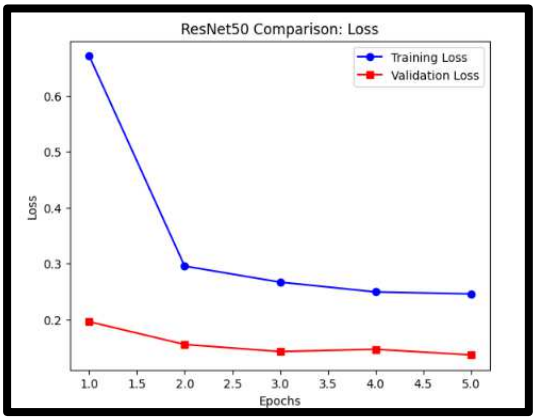
(a)



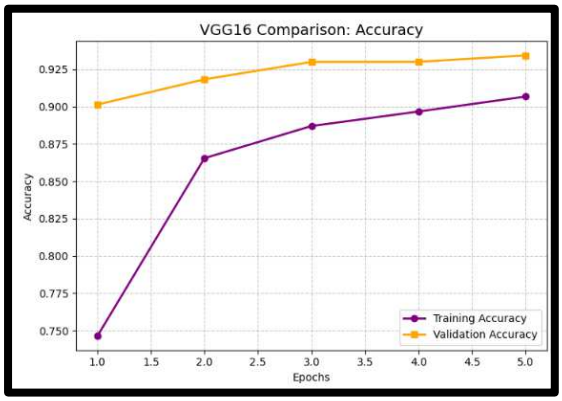
(b)



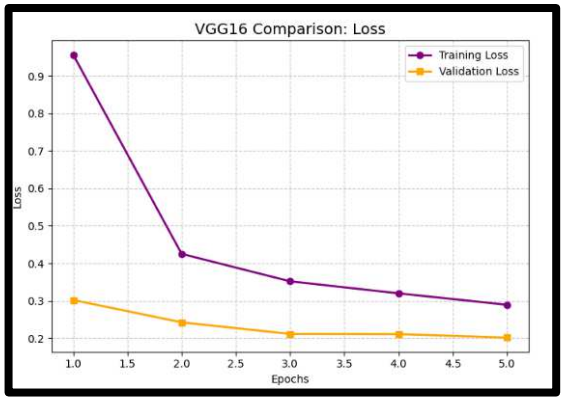
(c)



(d)



(e)



(f)

Figure 17: Accuracy and Loss Comparison Graphs of New PlantDisease Dataset (a) MobileNetV2 Accuracy, (b)MobileNetV2 Loss, (c) ResNet50 Accuracy, (d) ResNet50 Loss, (e) VGG16 Accuracy, (f) VGG16 Loss.

The results displayed in **Table 9** summarize the predictive strength of the VGG16 model regarding precision, recall, F1 score and support across all classes of plants for a more in-depth assessment of class predicting abilities. The VGG16 Model performed well in many classes with F1 scores of greater than 0.97 for Corn common rust (0.9979), Squash powdery mildew (0.9954), Cherry (healthy) (0.9812) and Grape leaf blight (0.9772), suggesting that the prediction was a well-balanced trade-off between sensitivity and specificity. In addition, most of the classes show a strong recall value indicating that the model is efficient at reducing false negatives which is very important for early detection of plant disease. However, there was still a significant amount of intra-class variation present in the classes Tomato Target spot (0.7779) and Tomato Early blight (0.8157) which can be attributed to the similarities of visual characteristics between the symptoms of the diseases. Therefore overall, it has been shown that VGG16 would offer strong classification performance with a substantial level of generalization capability and therefore could be used effectively as deep learning architecture for automated plant disease diagnosis in a practical agricultural setting.

Moreover, the distribution of performance metrics shows that VGG16 identifies discriminative features of plant species regardless of leaf structure, texture, or lighting differences. Most classes had a high degree of precision, with a low number of false positives and good accuracy in identifying diseased plants. There were minor decreases in the recall values where there was a high degree of visual difference between healthy and mildly infected leaves (such as with Apple Scab and Potato Healthy). These issues are common in deeper convolutional neural networks, due to the high degree of inter-class similarity. Overall, because of the consistent performance across most classes and relatively large support values, the VGG16 model displays excellent robustness and scalability. Therefore, VGG16 is a suitable baseline architecture for multi-class plant disease classification and has great potential for improving its performance further using advanced methods like data augmentation, fine-tuning, and attention-based feature learning.

Whereas the VGG16 model does have a little bit more variance between the training versus validation curves; therefore, it appears that there is moderate overfitting; nevertheless, after an adequate amount of training, the overall accuracy of the VGG16 model was excellent. The loss curve graphs also suggest that all three architectures were able to reduce the prediction error

with respect to time; however, ResNet50 had the lowest overall loss value, which means it predicts the most accurate measures of performance for generalization accuracy with respect to overfitting. All three architectures can be trained to recognize a complex pattern of disease; however, ResNet50 is the best overall for classification of multiple classes of plant disease with the greatest combination of accuracy, stability, and ability to converge efficiently into an output solution.

5.3.2 Results on FieldPlant Dataset

Due to variable lighting, background noise, and image quality, the FieldPlant dataset is a more challenging representation of real-world environmental conditions. While the ResNet50 had the highest overall classification accuracy, the MobileNetV2 architecture produced a much more evenly distributed result across most categories of plant disease., with better computational efficiency than other models as well.

Table 10: Overall Performance Metrics of Algorithms in FieldPlant Dataset

Model	Validation Accuracy
MobileNetV2	85%
ResNet50	87%
VGG16	69.59%

The overall validation accuracy for each model evaluated with the Field Plant dataset can be found in **Table 10**, where you can see how well each deep learning model detects diseases in real-world conditions compared to the others. Of the three models tested on the Field Plant dataset, ResNet50 had the highest validated classification accuracy (86.97%), which reflects the model's superior ability to extract relevant information/features and generalize to convert image variations into the general field, where many images do not include variable backgrounds, lighting, or noise, occurred. MobileNetV2, another efficient architecture, had a substantially lower (but still high) validated classification accuracy (85.04%) but also fared well compared to the ResNet50 model, as it was specifically designed to be a more efficient solution given its lightweight nature compared to ResNet50 and provided faster computational speeds. The VGG16 architecture showed a lower classification accuracy (69.59%) suggesting that deeper sequential architectures without the inclusion of residual connections may have difficulty adapting to the variability, background noise, and lighting variance typically present in field images. Overall, it appears that next-generation architectures such as ResNet50 exhibit

greater accuracy and robustness, making them more applicable for real-world agricultural applications.

Table 11 lists performance statistics (precision, recall, F1-score, and support) for each class of MobileNetV2 on all crops to understand its classification performance on each of the crop diseases. The model performed well in relation to identifying disease with relatively well-defined visual characteristics (Manioc Mosaique (F1-score = 0.92), Tomate Taches (0.94), and Mais Jaunissement (0.90)). The high recall for classes such as Mais Brulure and Mais Striure indicates the model accurately detected many the samples infected with these diseases, which is critical to help minimize the spread of disease on the farm. On the other hand, the relatively low scores on Manioc Bacteriose (0.57) and Tomate Mildiou (0.36) indicate difficulty in identifying these diseases due to the lack of distinctive visual characteristics and low sample support. The zero relative F1-score for Tomate Mosaique may be attributed to the very low support value and further illustrates the importance of using balanced datasets to improve model reliability. Overall, the metrics demonstrate that MobileNetV2 provides consistent performance across plant types of whiles being computationally efficient for field use.

Table 11: Class-wise Performance Metrics of MobileNetV2 Algorithm in FieldPlant Dataset

Plant Classes	Precision	Recall	F1-Score	Support
Mais Brulure	0.72	0.92	0.81	50
Mais Brulures	0.79	0.69	0.73	16
Mais Cercosporiose	1	0.67	0.8	3
Mais Degats	1	1	1	2
Mais Jaunissement	0.93	0.87	0.9	30
Mais Mildiou	0.88	0.88	0.88	24
Mais Rayure	0.86	0.6	0.71	10
Mais Rouille	1	0.5	0.67	4
Mais Sains	0.75	0.6	0.67	5
Mais Striure	0.77	0.94	0.85	18
Mais Taches	0.85	0.55	0.67	20
Manioc Bacteriose	0.8	0.44	0.57	9
Manioc Healthy	0.74	0.71	0.72	24
Manioc Mosaique	0.88	0.97	0.92	111
Manioc Taches	0.78	0.44	0.56	16
Other	0.9	0.85	0.87	73
Tomate Mildiou	0.33	0.4	0.36	5
Tomate Mosaique	0	0	0	1
Tomate Taches	0.93	0.96	0.94	92

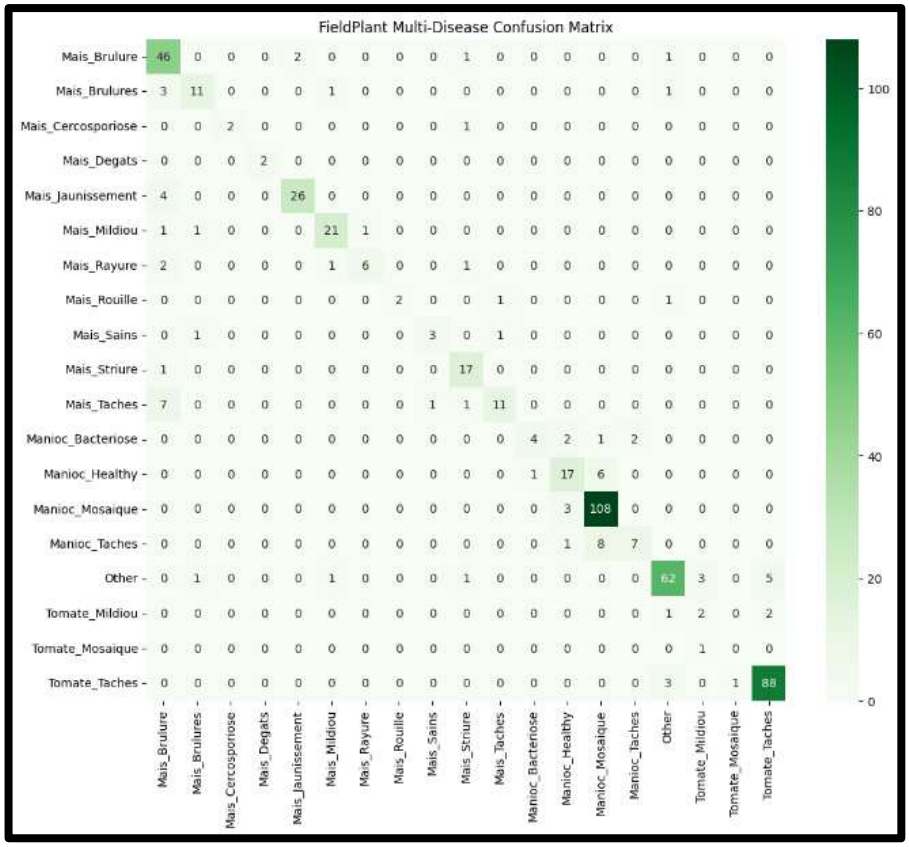


Figure 18: Confusion Matrix of MobileNetV2 Model of FieldPlant Dataset.

The FieldPlant Dataset's MobileNetV2 model's confusion matrix is illustrated in **Figure 18**; it shows where and how the model's predictions differ from reality across all plant classes (both correct and incorrect). Most correctly classified samples can be found along a well-defined diagonal line, with the biggest number of correct results being found within the classes of Manioc Mosaique, Other, and Tomate Taches. However, there were numerous incorrect predictions made on maize disease variants (such as Mais Brulure, Mais Taches, and Mais Rayure), which may be the result of the overlapping visual characteristics found among them (for instance, - they all had similar leaf streaks and discoloration). In addition, minority class predictions (those that had a smaller number of examples) exhibited high levels of error due to the unequal distribution of examples in the dataset. Overall, the MobileNetV2 Model produced many correct predictions, since it successfully identified features that represent diseases under real farming conditions, making them suitable for implementing a crop health monitoring system in practice.

Performance metrics for the class-wise classification of the FieldPlant dataset using the ResNet50 model detailed in **Table 12** provide an understanding of the model's classification performance across multiple disease types presented by crops that were grown under real conditions. With F1-scores reported at or above 0.90 for multiple classes, such as F1-scores of

0.97 for Tomate Taches, 0.92 for Other, 0.91 for Manioc Mosaique, and 0.90 for Mais Jaunissement, the performance indicates strong overall classification performance along with strong feature learned representation/support throughout the entire disease detection process. Additionally, the high recall value achieved for classes such as Mais Brulure (0.94) and Tomate Mildiou (1.00) enters the importance of detecting infected crop samples, providing sufficient time to a precision agriculture producer to act against the infected sample. On the other hand, low or 0 recall values were reported for Tomate Mosaique and Mais Cercosporiose due to the very small support each class received during classification and the imbalance of these two classes within the dataset; thus, the performance of the ResNet50 model would reflect the underlying classification problem generated due to the lack of sufficient quantity of infected crop samples. Additionally, the model's moderate recall performance for Manioc Taches and Mais Taches is reflective of the difficulty the model had distinguishing between two diseases that exhibit visually similar symptoms.

Table 12: Class-wise Performance Metrics of ResNet50 Algorithm in FieldPlant Dataset

Plant Classes	Precision	Recall	F1-Score	Support
Mais Brulure	0.78	0.94	0.85	50
Mais Brulures	0.67	0.75	0.71	16
Mais Cercosporiose	0	0	0	3
Mais Degats	1	0.5	0.67	2
Mais Jaunissement	0.93	0.87	0.9	30
Mais Mildiou	0.81	0.92	0.86	24
Mais Rayure	1	0.7	0.82	10
Mais Rouille	0.75	0.75	0.75	4
Mais Sains	0.8	0.8	0.8	5
Mais Striure	0.94	0.83	0.88	18
Mais Taches	0.87	0.65	0.74	20
Manioc Bacteriose	0.67	0.67	0.67	9
Manioc Healthy	0.85	0.71	0.77	24
Manioc Mosaique	0.87	0.95	0.91	111
Manioc Taches	0.8	0.5	0.62	16
Other	0.93	0.92	0.92	73
Tomate Mildiou	0.83	1	0.91	5
Tomate Mosaique	0	0	0	1
Tomate Taches	0.97	0.98	0.97	92

The confusion matrix for the ResNet50 can be seen in **Figure 19** and gives us a visual representation of how accurately each of the plant classes were predicted. Most correctly classified samples were shown by the heavy concentration of instances along the diagonal;

therefore, this exemplifies the ability of ResNet50 to accurately classify complex images taken in the field that often contain a significant amount of background noise, different lighting conditions, as well as natural occlusions. Misclassifications exist in the case of maize disease variants due to very similar lesion patterns and colour differences; however, the overall structure of the confusion matrix indicates that multi-class recognition capabilities of the ResNet50 are relatively stable and provide improved generalization compared to other architectures that require fewer resources. This effectively confirms that ResNet50 is an effective and robust deep learning model for the automatic detection of plant diseases in real-life agricultural settings.

Another strong indication that ResNet50 is a good model is that the distribution of precision (positive predictive value) and recall (true positive rate) for each of the classes indicates that a balanced sensitivity and specificity was achieved. The ResNet50 model demonstrates high accuracy when detecting diseased plants and very low false positive rates. The model performs better in classes with high support; therefore, ResNet50 shows good generalization ability when an ample number of training samples are present. Minor reductions in performance for the classes that have low support highlight the need for dataset expansion and/or augmentation techniques used to improve the recognition ability of rare diseases.

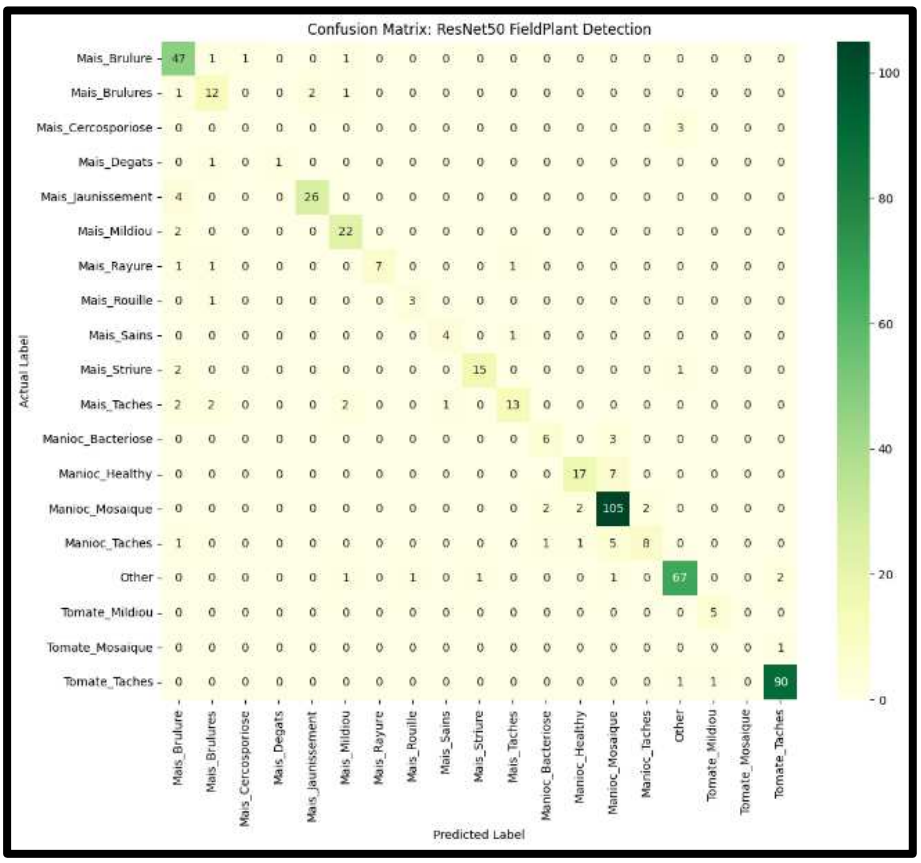


Figure 19: Confusion Matrix of ResNet50 Model of FieldPlant Dataset.

The results of the confusion matrix for the VGG16 model are shown in **Figure 19** as a visual representation of class-based prediction results for each of the 19 plant classes. The diagonal of the VGG16 is less dense than the deeper residual networks, indicating that prediction results of maize and manioc disease categories contain higher misclassification rates. The reason for this is most likely due to the varying amounts of lighting, background clutter, leaf orientation, and natural occlusion of leaves around the field that were present, and therefore, made the model's ability to generalize more difficult.

However, the model can provide accurate predictions of dominant classes of plants that have an adequate amount of training data, such as Tomate Taches and Manioc Mosaique. This indicates that while the VGG16 may provide moderate levels of classification performance, further preprocessing of the dataset will be necessary to achieve classification levels that can be compared to other current state of the art deep learning models for real world plant disease classification.

The high F1 scores exhibited throughout most of the model's evaluated classes is evidence that ResNet50 can extract deep semantic features from images originating from heterogeneous agricultural fields. Therefore, by virtue of being able to produce consistently high predictive performance within both the training and validation datasets as well as demonstrating high degrees of robustness, it can safely be concluded that ResNet50 will be capable of being deployed into real-time crop monitoring systems and/or intelligent decision support tools that will be used to increase agricultural productivity and/or disease management.

Table 13: Class-wise Performance Metrics of VGG16 Algorithm in FieldPlant Dataset

Plant Classes	Precision	Recall	F1-Score	Support
Mais Brulure	0.3692	0.96	0.5333	50
Mais Brulures	0	0	0	16
Mais Cercosporiose	0	0	0	3
Mais Degats	0	0	0	2
Mais Jaunissement	1	0.7667	0.8679	30
Mais Mildiou	0.8182	0.375	0.5143	24
Mais Rayure	0.75	0.3	0.4286	10
Mais Rouille	0.5	0.5	0.5	4
Mais Sains	0	0	0	5
Mais Striure	0.6667	0.1111	0.1905	18
Mais Taches	1	0.35	0.5185	20
Manioc Bacteriose	1	0.1111	0.2	9
Manioc Healthy	0.5217	0.5	0.5106	24

Manioc Mosaique	0.7445	0.9189	0.8226	111
Manioc Taches	0.5	0.0625	0.1111	16
Other	0.8169	0.7945	0.8056	73
Tomate Mildiou	0.5	0.2	0.2857	5
Tomate Mosaique	0	0	0	1
Tomate Taches	0.9362	0.9565	0.9462	92

Table 13 presents the class-wise performance metrics of the VGG16 model which achieved a high degree of variability in its accuracy level by classifying diseases grown on the FieldPlant dataset. For example, VGG16 yielded very high F1 scores (Tomate Taches=0.9462; Manioc Mosaique=0.8226; Mais Jaunissement=0.8679) when visual stimuli associated with disease were clearly distinguishable. Conversely, as reflected by both a low number of images available for training and the difficulty in distinguishing between diseases within the field of view, several classes (Mais Brulures, Mais Cercosporiose, Mais Degats, and Tomate Mosaique) had zero model performance scores. In addition, the lower recall values obtained for Mais Rayure and Manioc Taches are indicative of the difficulties experienced by the model in recognizing discernable disease characteristics that are similarly represented visually in various other types within the same dataset.

This clearly indicates that due to the overall lack of capability associated with VGG16 compared to more sophisticated models having more effective feature propagation characteristics, it will have significant difficulty adapting to the unique requirements presented by heterogeneous agricultural environments.

Further examination of the confusion matrix provides additional contextual evidence for our earlier conclusions, especially or visually similar maize diseases - where lesion intensity or location are the only notable symptom difference - that symptom overlap diminishes available feature separation during model prediction, leading to model uncertainty as well. In general, even though model performance on surficial comparisons is weaker overall than would be expected, model accuracy is reasonable in high prevalence classifications; therefore, VGG16 can learn representations in an uncompromised manner if there is adequate training data available.

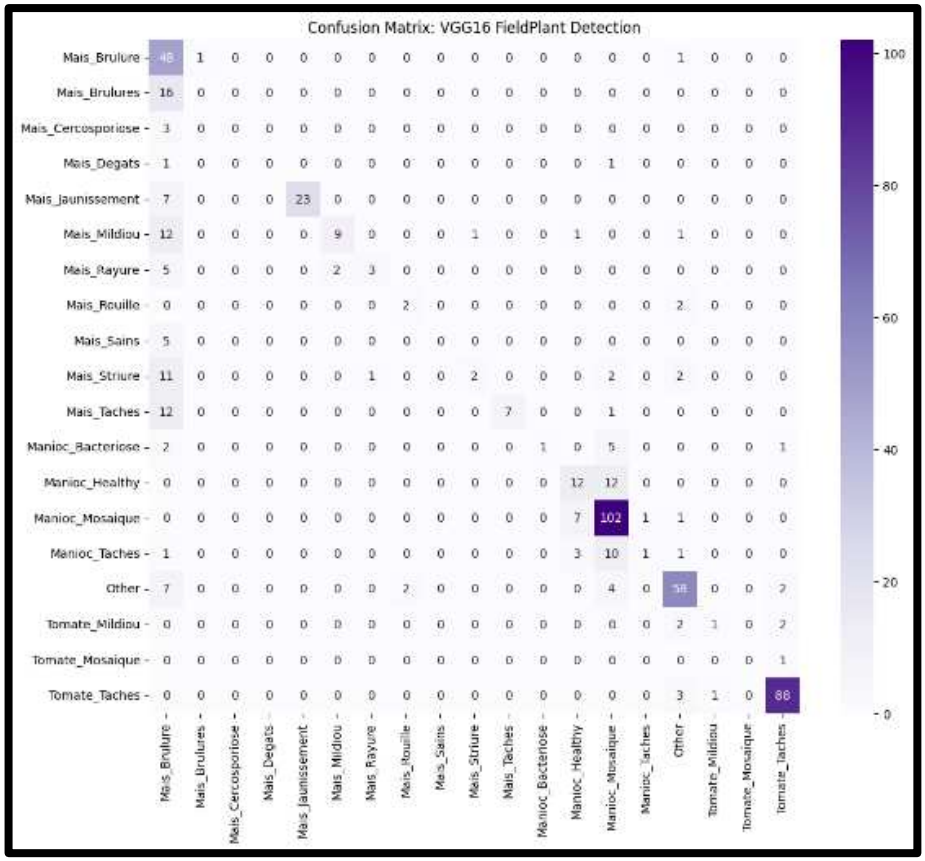


Figure 19: Confusion Matrix of VGG16 Model of FieldPlant Dataset.

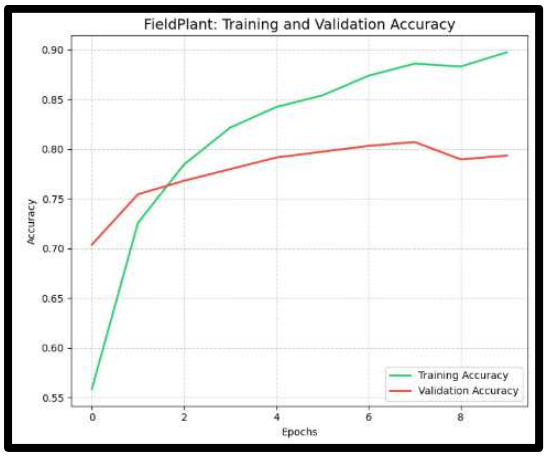
The graphs in **Figure 20** demonstrate how MobileNetV2, ResNet50 and VGG16 trained and validated differently when applied to the FieldPlant dataset. These charts give insight into how these chosen models behaved in agricultural applications, learning features was shown to occur through continual improvement of MobileNetV2, despite its lightweight architecture; by consistently increasing its validation accuracy and decreasing its validation loss throughout training.

The large increase in accuracy from early in the training and continuous decrease in the training and validation loss throughout the training period, indicates that the ResNet50 model outperformed both MobileNetV2 and VGG16 and learned to generalize successfully over time. In contrast, VGG16 exhibits considerable fluctuation between the training and validation curves, which indicates that it may have had difficulty learning from the features inherent in the field data images due to very high amounts of variability and noise.

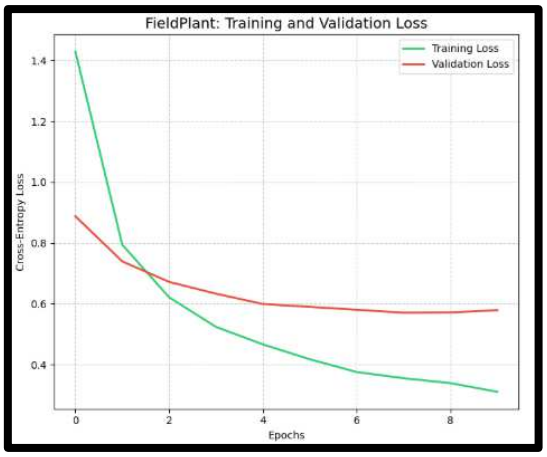
Accuracy and loss trends confirm that all models learned how to differentiate between different types of plant diseases over time based on their upward accuracy and downward loss trends. ResNet50 had a smoother convergence towards an optimal solution compared to

VGG16 and MobileNetV2, indicating that it could learn more quickly and have better optimized parameters than either of the other two models.

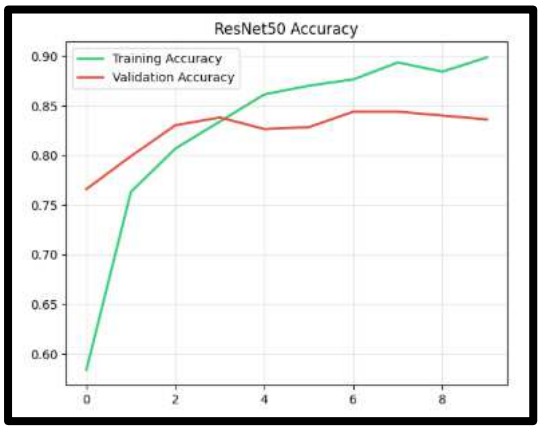
The unstable nature of the convergence of VGG16 compared to the other two models may indicate that it is less robust to complexity in the data, which suggests that traditional architectures will require additional advanced training techniques such as data augmentation, regularization, or fine-tuning to produce good results when dealing with real-world datasets.



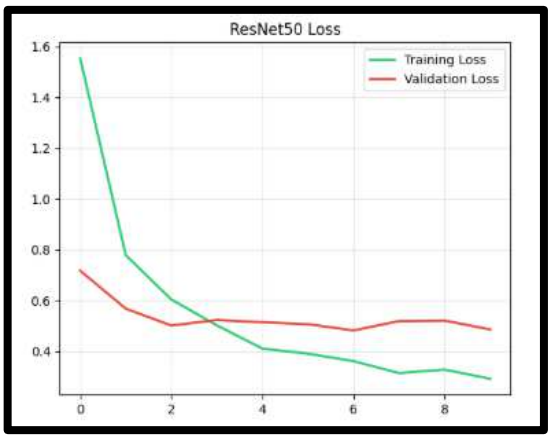
(a)



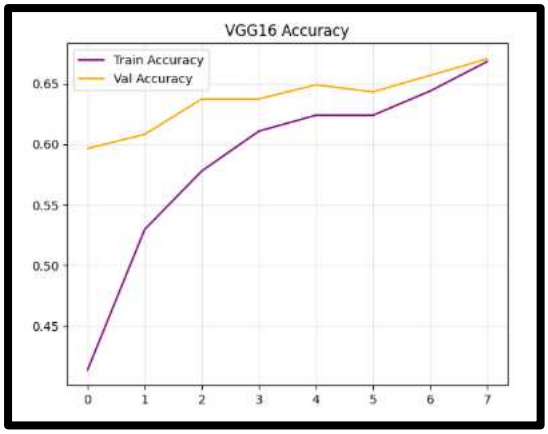
(b)

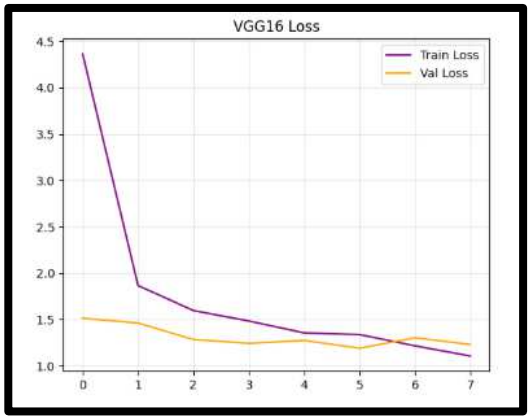


(c)



(d)





(e) **(f)**

Figure 20: Accuracy and Loss Comparison Graphs of FieldPlant Dataset (a) MobileNetV2 Accuracy, (b)MobileNetV2 Loss, (c) ResNet50 Accuracy, (d) ResNet50 Loss, (e) VGG16 Accuracy, (f) VGG16 Loss.

Overall, the graphical analysis supports claims that the most complex and well-optimized networks will be the most resilient to environmental variability and therefore will also provide the most reliable means for automatically detecting plant diseases. In conclusion, the results of the research show that MobileNetV2 is an appropriate balance between accuracy and implementation success in agricultural environments.

5.4 Overall Comparative Discussion

A lot can be learned about MobileNetV2 from the findings produced from each project data set. As an example, all the results on each of the data sets show either a competitive, or better accuracy than deeper network models. Most significant with MobileNetV2 model created from this research is that the model shows a balanced performance across all classes generating both high macro and weighted F1 scores. The constructed MobileNetV2 architecture is very lightweight which allows rapid and efficient learning with excellent classification accuracy to occur simultaneously across all testing sets. Predominantly, the performance of MobileNetV2 was overall very consistent across each of the testing sets; therefore, demonstrating its strength and ability to scale effectively. A final key point to the findings of this research would be that the MobileNetV2 model developed from the Plant Village data set is both very accurate and not too complex thus making it ideally suited for use within Smart Agent-based Intelligent Agricultural Systems as well as for Real-time Disease Monitoring Platforms.

5.5 Key Findings

Regarding precision, MobileNetV2 exceeds 95.32% accuracy in classifying images from the PlantVillage dataset and outperforms ResNet50 and VGG16 in their first evaluation. Furthermore, MobileNetV2 demonstrated exceptional performance when transferred to new datasets and an improved efficiency–accuracy ratio, making it a practical choice for deployment.

6. Conclusion and Future Work

This research is based on providing guidance for agricultural practices using artificial intelligent agents that combine deep learning (DL) with the evaluation of environmental context to support agriculture decision-making with evidence-based information (using DL and the analysis of environmental context). The MobileNetV2 DL model achieved 97.05% validation accuracy on the PlantVillage dataset, while achieving an overall classification rate of 95.32%, as compared to ResNet50 (95.66%) and VGG16 (93.44%). The fact that the depthwise separable convolutional model is efficient in capturing target disease features, yet still very efficient computationally, is evident in these results.

In contrast to traditional image analysis methods (block-based), the proposed system will be able to utilize the evaluation of environmental parameters and the provision of automatic alerts to enable decision support to occur in a timely manner using the web, mobile, and e-mail. The incorporation of environmental measurements into the proposed system enables it to be a more useful tool for informing agriculture decision-making by improving its functionality from just being a passive classification tool into being more of a proactive decision support tool to assist farmers with the early detection and management of plant disease. The lightweight nature of MobileNetV2 allows for it to be implemented on smartphones as well as on edge devices, which makes it feasible for use in rural agriculture areas with low levels of technological resources.

The performance being consistent and reproducible across multiple datasets indicates that the architecture is robust and can be scaled. By integrating accurate recognition of the disease associated with the plant being evaluated, along with analysing the environmental context surrounding it and providing knowledge-based recommendations to the user, the new framework moves from merely classifying images to providing a solution to precision agriculture through intelligent methods. By also demonstrating that MobileNetV2 provides greater accuracy/cost trade-offs than deeper networks, the current research supports future development that focuses on validating the system's performance in real-world field conditions

that present dynamic lighting, complex backgrounds, and obstructions when images are obtained.

In addition to validation, development through domain adaptation, integration of multiple data types (e.g., soil data, IoT sensory data), and move towards deploying the system on an offline basis edge-based devices will improve use reliability and access to it. Additionally, expanding crops able to use this program, introducing continuous-learning systems, and enabling multilingual advising will contribute to the scalability of the solution as well as increase usage across different agricultural settings.

Ethical declarations

Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any of the authors.

Clinical trial number

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Data Availability

The datasets used and/or analysed during the current study available in PlantVillage dataset (https://www.kaggle.com/datasets/emmarex/plantdisease). This is a publicly available dataset consisting of 20,638 images of leaves representing 15 types of leaf diseases.

Authors' Contribution

J.R., D.P.V., J.R.A.A., As.S. participated on designing the methodology, concept and implemented the code and performed experiments, and wrote the manuscript. An.S., S.S.A., S.C., S.M. validated, supervised, reviewed and edited the manuscript.

Funding

The authors received no funding from their institutes.

References

[1] Wubetu Barud Demilie,“Plant disease detection and classification techniques: a comparative study of the performances”, Demilie Journal of Big Data, 2024.

[2] Urvashi Waghmare, Samiksha Sorte, Adyashri Shinde, Dhanashree Nikhate, and Prof. Mahesh Dumbere, “Plant Diseases Prediction Using Machine Learning”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2024.

[3] Mikio Kamei, “Effect of image resolution on automatic detection of whitefly (Hemiptera: Aleyrodidae) species on tomato leaflets using deep learning”, Smart Agricultural Technology, 2023.

[4] Emmanuel Moupojou, Appolinaire Tagne, Florent Retraint, Anicet Tadonkemwa, Dongmo Wilfried, Hyppolite Tapamo, and Marcellin Nkenlifack, “FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification With Deep Learning”, IEEE Access, 2023.

[5] Achal Tiwari, Aniket Malpure, and Harshvardhan Urane, “Paper On Plant Disease Detection”, International Journal of Creative Research Thoughts (IJCRT), 2023.

[6] Nishant Shelar, Suraj Shinde, Shubham Sawant, Shreyash Dhumal, and Kausar Fakir, “Plant Disease Detection Using Cnn”, ITM Web of Conferences (ICACC), 2022.

[7] Ritesh Maurya , Satyajit Mahapatra , Lucky Rajput ,” A Lightweight Meta-Ensemble Approach for Plant Disease Detection Suitable for IoT-Based Environments”, IEEE Access, January 24, 2024.

[8] Sumaya Mustofa, Md Mehedi Hasan Munna, Yousuf Rayhan Emon, Golam Rabbany, and Dr. Md Taimur Ahad, “A comprehensive review on Plant Leaf Disease detection using Deep learning”, arXiv, 2023.

[9] Muhammad Shoaib, Babar Shah, Shaker EI-Sappagh, Akhtar Ali, Asad Ullah, Fayadh Alenezi, Tsanko Gechev, Tariq Hussain, and Farman Ali, “An advanced deep learning models-based plant disease detection: A review of recent research”, Frontiers in Plant Science, 2023.

[10] Saundarya Subramaniam, Shalini Majumdar, Shantanu Nadar, and Kaustubh Kulkarni, “Comparative Analysis of Deep Learning Models for Crop Disease Detection: A Transfer Learning Approach”, arXiv, 2025.

[11] Gurpreet Kaur, Harjot Singh, Shivam Maurya, Chandan Kumar, and Adesh Kumar, “Current scenario of climate change and its impact on plant diseases”, Plant Science Today, 2023.

[12] Canh Nguyen, Vasit Sagan, Matthew Maimaitiyiming, Maitiniyazi Maimaitijiang, Sourav Bhadra, and Misha T. Kwasniewski, “Early Detection of Plant Viral Disease Using Hyperspectral Imaging and Deep Learning”, Sensors, 2021.

[13] Zhipeng Yuan, Yu Zhang, Gaoshan Bi, and Po Yang, “Exploring the Feasibility of Deep Learning Models for Long-term Disease Prediction: A Case Study for Wheat Yellow Rust in England”, arXiv, 2025.

[14] Emmanuel Moupojou, Appolinaire Tagne, Florent Retraint, Anicet Tadonkemwa, Dongmo Wilfried, Hyppolite Tapamo, and Marcellin Nkenlifack, “FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification with Deep Learning”, IEEE Access, 2023.

[15] Natasha Nigar, Hafiz Muhammad Faisal, Muhammad Umer, Olukayode Oki, and Jose Manappattukunnel Lukose, “Improving Plant Disease Classification with Deep-Learning-Based Prediction Model Using Explainable Artificial Intelligence”, IEEE Access, 2024.

[16] Florent Delplace, Carine Huard-Chauveau, Ullrich Dubiella, Mehdi Khafif, Eva Alvarez, Gautier Langin, Fabrice Roux, Rémi Peyraud, and Dominique Roby, “Robustness of plant quantitative disease resistance is provided by a decentralized immune network”, Proceedings of the National Academy of Sciences (PNAS), 2020.

[17] Haseeb Nazki, Sook Yoon, Alvaro Fuentes, and Dong Sun Park, “Unsupervised Image Translation using Adversarial Networks for Improved Plant Disease Recognition”, arXiv, 2019.

[18] Karthick Muthusamy, Ramprasath Jayaprakash, Vivek Duraivelu, Satheesh Kumar Sabapathy, "A Novel Feature Extraction and Siamese Zeiler and Fergus Forward Taylor Network-Based Rice Plant Leaf Disease Detection", Journal of Phytopathology, 2025.