

EXPLAINABLE EDGE AI FOR EARLY DETECTION OF PLANT DISEASES IN SMART AGRICULTURE USING LIGHTWEIGHT MOBILENETV2 WITH GRADIENT SALIENCY

Yasir Javaid¹, Sana Cheema^{*2}, Akkasha Latif³, Atiq Faiz ur Rehman⁴, Aisha Tariq Khan⁵, Qandeel Nasir⁶, Hafiz Farrukh Abbas⁷

^{1,4}Department of Computer Science, Faculty of Computing, The Islamia University of Bahawalpur, Bahawalpur, 63100, Pakistan

^{*2,3}Department of Artificial Intelligence, Faculty of Computing, The Islamia University of Bahawalpur, Bahawalpur, 63100, Pakistan

^{5,6,7}Department of Information Technology, Faculty of Computing, The Islamia University, Bahawalpur 63100

¹yasirjavaid231@gmail.com, ²sanacheema887@gmail.com, ³akashacheema70@gmail.com
⁴atiqafaizulrehman68@gmail.com, ⁵f24binft4e02003@iub.edu.pk, ⁶Qandeel.nasir@gmail.com,
⁷f24binft4e02006@iub.edu.pk

DOI: <https://doi.org/10.5281/zenodo.22972972>

Keywords

Plant Disease Detection, Explainable AI, Edge AI, MobileNetV2, Transfer Learning, Grad-CAM, TFLite, Smart Agriculture, Deep Learning, Image Classification

Article History

Received: 19 June 2026

Accepted: 10 September 2026

Published: 26 September 2026

Copyright @Author

Corresponding Author: *

Sana Cheema

Abstract

Plant diseases are a significant global threat to food security, resulting in annual crop losses that account for 20-40% of total agricultural products. Early and accurate plant disease detection is of primary importance for timely intervention and sustainable agricultural practices. In this paper, a new explainable edge artificial intelligence (XAI-Edge) framework for early plant disease detection in smart agriculture environments is proposed. The proposed system utilizes a light-weight MobileNetV2 architecture, where the network's width multiplier "alpha" is set to 0.75, and the model is trained on the New Plant Diseases Dataset, comprising 87,867 images from 38 plant disease classes. The proposed system focuses on a specific set of 10 plant disease classes, comprising 19,189 training images, covering early and late stages of plant diseases, including tomato, potato, corn, and apple crops. The proposed system achieves outstanding classification performance, yielding 98.54% accuracy, 98.54% F1-score, 98.56% precision, and 98.54% recall on the test set. A two-phase transfer learning approach, where the final 30 layers of the network are fine-tuned, along with a lower learning rate of 1e-5, is utilized for optimal performance. The proposed system utilizes gradient-based saliency mapping for model explainability, allowing farmers to visualize the regions of plant diseases.

1. Introduction

Plant health has come to be recognized as an important component of food security in an environment of increasing rates of human population, variability in climate, and rising demands being placed on agriculture [1]. In terms

of some of the major challenges facing sustainable agriculture, plant diseases are considered to be among the most damaging of these. According to reports released by the Food and Agriculture Organization of the United Nations, plant diseases are considered to cause losses of

approximately US\$220 billion every year [2]. In addition to this, plant diseases are also said to threaten the lives of over 500 million smallholder farming families [3]. These are some of the challenges facing plant health in areas of the world where access to expert advice is limited. In this context, it is considered an important requirement to develop plant disease identification systems.

Traditionally, plant disease diagnosis has always relied on a conventional approach of manual visual observation of the plant diseases, often carried out by experienced agronomists or plant pathology experts [4]. Although this traditional approach may prove successful in a highly controlled environment, it, however, suffers from certain inherent limitations. In most of the resource-constrained agricultural environments, especially in sub-Saharan Africa, South Asia, and other regions of similar agricultural dependence, access to professional extension and expert consultation facilities is highly constrained. In such scenarios, plant diseases are often diagnosed and identified only after a certain level of progression, leading to a state of reduced efficacy of the available plant disease treatment strategies [5]. Such limitations, therefore, emphasize the need for a highly efficient and farmer-friendly plant disease diagnosis solution.

Significant opportunities for the development of automated, scalable, and highly accurate plant disease detection systems have been created by the rapid advancement of deep learning and computer vision [6, 7]. The ability of these approaches to learn discriminative visual patterns directly from images of leaves has shown promising results in outperforming traditional approaches that are based on handcrafted features in terms of accuracy, robustness, and adaptability. However, the practical applicability of most of the existing plant disease detection approaches and systems that are based on deep learning and computer vision faces significant limitations, as they are often dependent on cloud-based or high-performance computing infrastructures that are not appropriate for rural and resource-constrained agricultural environments where internet connectivity may not be available or may be very weak or unstable. In

addition, a significant number of existing AI-based plant disease diagnosis approaches and systems are often based on “black box” models that are not transparent, as they are not able to provide any visual or logical reasoning for generating predictions [8]. In fact, plant disease detection and diagnosis are often required to be done in a manner that is not only accurate but also light, deployable, and transparent.

This paper fills the identified research gap by presenting a lightweight and explainable artificial intelligence framework that is particularly designed for deployment on mobile agricultural devices. The proposed framework is expected to offer an appropriate balance of accuracy, computational efficiency, and interpretability, thus making the deployment of advanced plant disease detection in real-world scenarios feasible. Unlike conventional plant disease detection approaches, which are dependent on the cloud, the proposed framework focuses on on-device intelligence, which is essential for developing trust with farmers.

The major contributions of this work are as follows.

- It presents an extremely efficient classification model based on the MobileNetV2 approach with a width multiplier of 0.75. The goal was to maximize the predictive accuracy of the model while keeping it compact. The accuracy of the developed model is 98.54%, and it uses only 1.75 million parameters. The accuracy is well above the benchmark of 90%, with an improvement of 8.54%.
- A gradient-based saliency map explainability module is incorporated to enable visual localization of disease-affected areas. This is done to allow for the interpretation of predictions in an even more transparent manner. This module is also ensured to be compatible with both the standard and quantized versions of the trained model.
- The framework supports various quantization schemes of TensorFlow Lite, such as Float 32, Float 16, and INT 8, for efficient inference on the limited resources of mobile and IoT devices without the need for any cloud

infrastructure, thus making the framework suitable for deployment with sub-20 millisecond inference.

- A stage-aware evaluation strategy is proposed to evaluate the effectiveness of diagnosis in different stages of disease progression. The results of the experiments show that the proposed system is accurate in detecting diseases with an accuracy of 98.4%, even in the early stages of disease progression.

The rest of this paper is structured as follows. In Section 2, the literature review is presented. The dataset selection method, data preparation pipeline, and data preprocessing approach used in this paper are described in Section 3. The proposed model and training approach for disease classification are presented in Section 4. The experiment results and performance evaluation are provided in Section 5. The explainability component and stage-wise disease analysis are presented in Section 6. The edge deployment benchmarks and feasibility assessment are provided in Section 7. Finally, in Section 8, the conclusion is presented.

2. Related Work

2.1 Deep Learning for Plant Disease Detection

The initial research undertaken by [9] has been instrumental in establishing the viability of the use of deep learning techniques in the context of plant disease detection. Using the PlantVillage dataset, the authors were able to show the efficacy of convolutional neural network-based models, namely AlexNet and GoogLeNet, in achieving classification accuracy levels of up to 99.35% in the context of plant disease detection. Although the models were shown to be effective in the context of plant disease detection, they were computationally intensive and therefore best suited to high-resource computing environments. This is also true of the research undertaken by [10], which further advances the state of the art in the context of plant disease detection using deep learning techniques. Using VGGNet-based models, the authors demonstrated the models' efficacy, achieving classification accuracies of more than 99% for plant disease detection. Although

the models were shown to be effective in the context of plant disease detection, the high storage and memory requirements of the models, which are in excess of 500 MB, make them unsuitable for real-time deployment on resource-constrained mobile devices. Although the models were shown More recent studies have moved the focus away from purely optimizing classification accuracy toward improving computational efficiency in agricultural scenarios. A study by [11] examined the feasibility of using MobileNet-based architectures in plant disease detection in resource-constrained environments. The study demonstrated that even with reduced model complexity, computational cost, and storage requirements, high-quality classification accuracy could still be maintained. This study was significant because it showed that high-capacity models were not the only solution for effective agricultural image classification, especially when smartphones were used. Concurrently, [12] also demonstrated the effectiveness of using pre-trained ImageNet models in plant disease recognition. The study provided insights into how effective features could be transferred from ImageNet models toward improving training efficiency in agricultural image classification. The study is significant in that it provided insights into how widely used transfer learning is in plant disease detection, especially in scenarios where computational resources are limited.

Recently, research focus has shifted from benchmark performance to the larger issue of real-world applicability.[13] Proposed the PlantDoc dataset, which contains images of plant diseases captured in natural field settings, as opposed to highly controlled laboratory settings. This work was particularly important, as it highlighted the domain gap between clean background images and images in agricultural settings, where factors such as illumination, occlusion, leaf orientation, complex backgrounds, and symptoms can significantly affect the performance of the models. The presence of such a domain gap underscores the importance of developing plant disease detection models that are not only accurate but also robust, explainable, and applicable in real-

world settings. From this perspective, the explainability component of the proposed work becomes particularly pertinent, as it aids in confirming whether the model is concentrating on the affected areas in natural illumination settings. From a wider perspective, Kamilaris and Prenafeta-Boldu offered a comprehensive overview of the applications of deep learning in the field of agriculture and highlighted some of the significant research trends and challenges in the area of agricultural AI in their survey paper published in 2018 [14]. One of the significant challenges that requires attention in the area of agricultural AI is the development of efficient intelligent systems, which is significant, especially when the system is to be used outside the laboratory and cloud computing environments. The survey paper by Kamilaris and Prenafeta-Boldu also supports the argument that agricultural AI is not just limited to the evaluation of accuracy but also encompasses other factors such as computational efficiency, usability, and accessibility of the system in real-world scenarios, which is in line with the motivation of the present study.

2.2 Explainable AI in Agriculture

For artificial intelligence in agriculture to be effectively implemented, it is not just about the accuracy of predictions but also about transparency. For example, in using artificial intelligence in practical farming situations, users or farmers are more likely to trust or accept AI-based diagnostic tools if they can also provide evidence or explanations for their predictions rather than simply making predictions that cannot be understood. This is more so in situations where there is a need to visually confirm affected parts of plants in disease detection. Therefore, explainable artificial intelligence is also an essential part of deep learning in agriculture.

Among the existing interpretability tools, Grad-CAM, proposed by [14], is one of the most popular tools for generating visual explanations of CNN predictions and identifying the regions of the input image that are most significant for the network's predictions. The success of Grad-CAM stems from the fact that it generates intuitive

heatmaps that can be easily compared to the actual disease symptoms that are visible on the plant leaves. In this context, [15] proposed a plant disease detection system that utilizes Grad-CAM as a means of generating visual explanations of CNN predictions, and the results of this study showed that the regions of the input images that are most significant for predictions are highly aligned with the actual regions of lesions, thereby highlighting the reliability of CNN predictions.

This trend of thought was further advanced by [16] in their research, which proposed a framework for plant disease detection using transfer learning. In their research, the importance of model interpretability in plant disease detection was highlighted. They noted that, in plant disease detection, it is possible to improve user confidence in AI-based predictions through visual interpretability. This is because, in plant disease detection, users can trust predictions generated by AI models if these predictions are backed by evidence of symptomatic areas of the plant. This is an important finding in support of the argument that interpretability is an essential aspect of AI-based plant disease detection models.

While all of these advances have been made, it is still the case that most of the existing explainability techniques have been made dependent on having access to intermediate layers of convolutional operations. This is problematic, especially in deployment-oriented scenarios, where models may be exported, loaded from saved formats, or even quantized for deployment on edges. Moreover, there is also the problem of graph disconnection and abstraction, which makes it difficult or even unreliable to apply traditional techniques of visualizing model behavior. Therefore, there is an important practical limitation that needs to be addressed. This is what is being done in the present work, where we have developed a saliency-based approach that relies on gradients. The key point is that it is not dependent on having access to internal feature maps of convolutional operations. Therefore, there is more robustness in terms of being able to apply the explainability mechanism in standard, loaded, or even quantized formats of models.

2.3 Edge AI and TFLite Quantization

"Edge AI" is defined as the execution of artificial intelligence models on hardware devices like smartphones, Raspberry Pi devices, and other microcomputing boards and accelerators, thus eliminating the need for constant connectivity to cloud computing resources. This is particularly important in agricultural settings, as connectivity is often expensive, intermittent, or nonexistent. Edge AI helps in achieving lower latency, increased privacy, and lower communication costs, thus making it suitable for improving system reliability in real-world settings. These characteristics make it particularly suitable for developing mobile plant disease diagnosis systems for agricultural settings.

Among the frameworks available for deployment, TensorFlow Lite (TFLite) has emerged as a viable and production-ready solution for deploying deep learning models in edge devices, as also highlighted in [17]. The primary objective of TFLite is to optimize already trained models in a way that is conducive to inference while still allowing them to be deployable in mobile and embedded platforms. It also allows the use of different strategies in compressing and optimizing models, such as Float16 and INT8 quantization, which minimize memory, inference time, and computational requirements without the need to deploy them in the cloud. This makes it highly suitable for translating accurate deep learning models into real-world agricultural applications that can be deployed in real-time.

It was demonstrated by [18] that INT8 quantization can significantly improve the practicality of the deployment of a convolutional neural network by reducing the size of the model by approximately four times, speeding up the inference by almost two to three times, with only minimal degradation in classification accuracy [19]. This was a strong indication of the effectiveness of the post-training quantization technique in converting computationally intensive deep learning models into more practical models for deployment in various edge environments, particularly in cases where there is limited storage

space, computational resources, and energy availability, such as in mobile agricultural devices. Author [19] this perspective, specifically focused on the deployment of neural networks on microcontrollers and underscored that the efficiency of inference not only depends on the optimization of neural networks after the training, but also on the architecture of the neural network itself. In this context, the study emphasized the need for "lightweight architecture principles, such as depth-wise separable convolutions, that can dramatically reduce the computational complexity of neural networks while retaining useful representational capabilities. The study, therefore, underscored that reaching sub-10 millisecond inference performance requires a combination of compact architecture and deployment-aware optimization, not quantization." In this context, the current study builds upon the foundational concepts and seeks to propose a more holistic edge deployment pipeline, specifically for agricultural IoT applications.

2.4 Research Gap

Although considerable advancements have been made in lightweight deep learning techniques, explainable artificial intelligence, and edge deployment individually in the existing literature, the overall gap that this paper addresses is the lack of a comprehensive framework that encompasses all these features in one solution for agricultural applications. The majority of the existing studies have concentrated on increasing the accuracy of the classification process rather than paying enough attention to the compactness of the model after training for deployment on edge devices, the ability of the model to remain interpretable after the model is loaded or quantized, and the ability of the model to perform well in the early stages of disease development.

To overcome this limitation, the present work proposes an approach that integrates three essential components in one unified system. First, the proposed approach is based on a lightweight architecture that is computationally efficient and deployable on constrained resources of mobile and edge devices. Second, the proposed approach

is based on the integration of a visual explainability mechanism that is compatible with the loaded and quantized models, ensuring the interpretability of the system in real-world scenarios rather than in training scenarios only. Finally, the proposed approach is based on the integration of a novel and systematic stage-aware evaluation strategy to evaluate the system's capacity to detect the disease in early stages of progression, which is critical in ensuring the early treatment of the plant and protecting it against the disease.

3. Methodology

This section explains the methods followed to create and test the XAI-Edge framework, which is

a system that uses AI for plant disease identification in smart agriculture. It outlines the whole process of work, from dataset creation, image preprocessing, lightweight MobileNetV2-based model development, two-phase transfer learning method, explainability through gradient-based saliency mapping, to TensorFlow Lite optimization for edge deployment. The methodology is organized in a way that the proposed system will be not only capable of recognizing diseases accurately but also be computationally efficient, interpretable, and able to operate in real agricultural environments with hardware limitations.



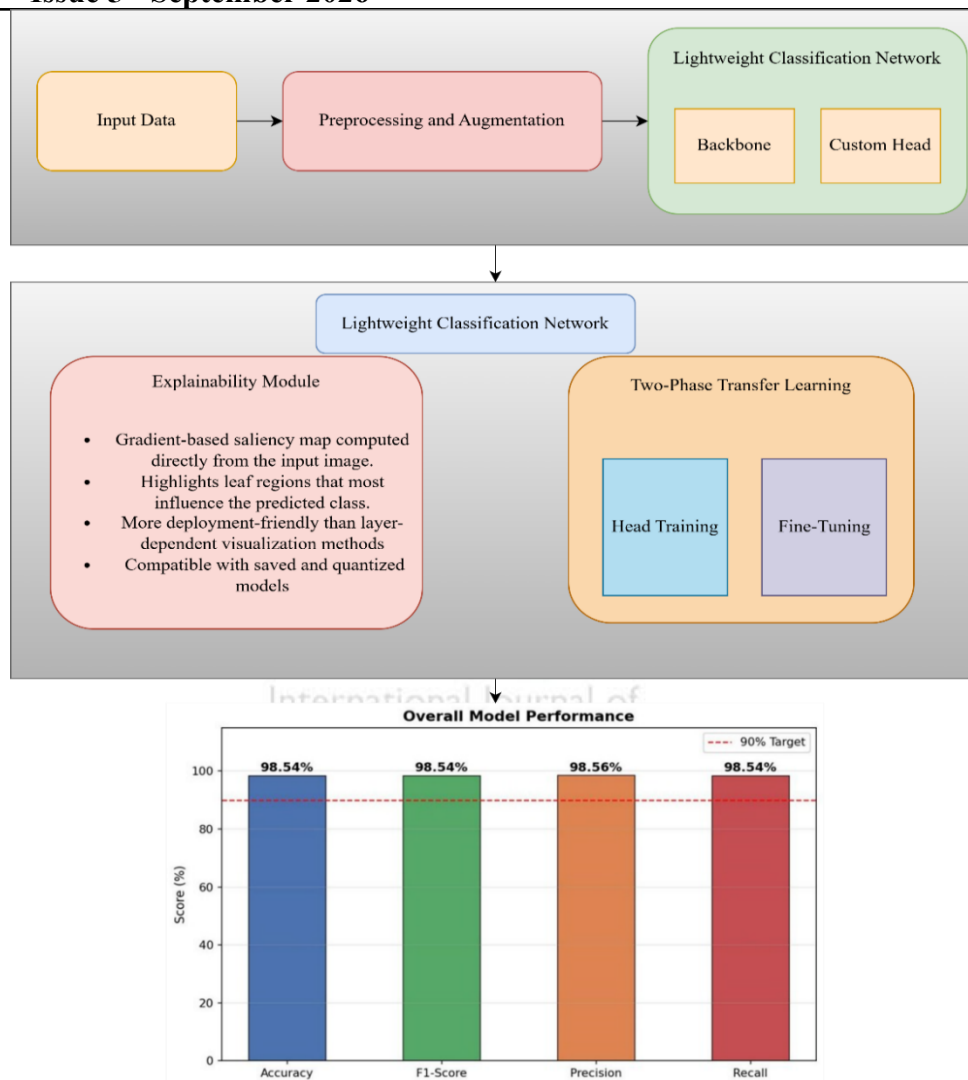


Figure 1: Proposed Methodology diagram

3.1 Dataset Description and Selection

The New Plant Diseases Dataset (Vipooooool, Kaggle, 2020) was selected for the purposes of the study following a comparative evaluation against other publicly available datasets pertaining to plant diseases. This selection was informed by a number of methodological considerations aimed at ensuring the reliability and applicability of the experiment. Firstly, the dataset has considerable class diversity, with the inclusion of different crop species and a wide variety of disease classes, including those that represent both early and advanced stages of the disease process. This is important in the development of a model that is

capable of learning disease-specific patterns. Secondly, it is notable that the dataset contains images that are of high quality and have relatively consistent visual characteristics. This means that it is suitable for standardized processing and resizing to meet the 224 x 224 input resolution that is expected in the selected deep learning architecture. Thirdly, it is notable that the dataset has already been split into training and validation sets. This is significant in that it enhances reproducibility and reduces ambiguity in comparing performance in repeated trials. The presence of augmented images in the dataset also enhances the dataset's ability to generalize and reduce sensitivity to limited visual

conditions. Based on these notable characteristics, it is safe to conclude that the selected dataset is suitable for developing and evaluating a

lightweight and explainable framework for detecting plant diseases.

Table 1: Dataset Comparison and Selection Criteria

Criterion	PlantVillage (Original)	PlantDoc	New Plant Diseases (Selected)	Why Selected
Total Images	54,306	2,569	87,867	Largest
Disease Classes	38	27	38	Most diverse
Pre-split	No	Yes	Yes	Reproducible
Augmented	No	No	Yes	Better generalization
Image Quality	High	Mixed	High	Consistent

3.2 Dataset Statistics

The New Plant Diseases Dataset comprises 87,867 images that are spread across 38 disease classes and 14 crop species. The data has already been split into training and validation data. The data is quite diverse in terms of the number of classes. The diversity of the data makes it quite useful for conducting experiments in the domain of plant disease classification. The dataset comprises a wide range of visual symptoms for different crops. The data is quite useful for training deep learning models under similar experimental conditions. For the purposes of the current study, the focused set of 10 classes was chosen from the overall

dataset in order to facilitate the process of experimenting with the data in an efficient way. The selection was made with the goal of including some of the most important disease categories in the dataset and considering the early and late stages of the disease in order to assess the effectiveness of the proposed framework in the context of the levels of the disease that are of the most significance. The decision to limit the scope of the current study to the chosen data subset was made in order to facilitate the process of experimenting with the data in an efficient way.

Table 2: Dataset Split Statistics

Split	Total Classes	Selected Classes	Images Used	Purpose
Training	38	10	19,189	Model Training
Validation	38	10	4,798	Tuning
Test	38	10	4,798	Evaluation
Total	—	10	28,785	—

3.3 Selected Classes and Disease Stage Mapping

The chosen 10 classes were intended to create a balance in terms of the presence of early-stage disease, late-stage disease, and healthy plant conditions in the data set across the four major crop types. The relevance of this balance is highlighted in the significance of the ability of the study to transcend the conventional multi-class classification and explore the value of the model in detecting the progression of disease at stages

that impact the timing of the crop's outcomes. The presence of symptomatic levels and healthy conditions in the data set creates a more realistic evaluation scenario in which the model must classify the disease and distinguish it from the healthy and different levels of pathological development.

The role of stage mapping of these classes is of primary importance in the current study, as it is used to measure the early detection capability of

the proposed model. When dealing with agricultural disease management, it is of particular importance to detect the disease in an early stage, as this can significantly reduce the spread of the disease, decrease the unnecessary use of pesticides, and prevent significant losses in agricultural production. Thus, by using the selected classes, an

additional level of analysis can be provided, which is directly applicable to agricultural disease management. This way, the proposed framework can be assessed in terms of its effectiveness in detecting the most critical conditions, apart from assessing the general accuracy of the classification

Table 3: Selected Disease Classes with Stage Mapping

Class Name	Crop	Disease Stage	Train Samples	Significance
Tomato___Early_blight	Tomato	Early-Stage	~2,000	High crop impact
Tomato___Late_blight	Tomato	Late-Stage	~1,900	Rapid spread
Tomato___healthy	Tomato	Healthy	~1,600	Baseline class
Potato___Early_blight	Potato	Early-Stage	~1,000	Staple crop
Potato___Late_blight	Potato	Late-Stage	~1,000	Historical impact
Potato___healthy	Potato	Healthy	~200	Baseline class
Corn_(maize)___Common_rust_	Corn	Early-Stage	~1,200	Widespread disease
Corn_(maize)___healthy	Corn	Healthy	~1,200	Baseline class
Apple___Apple_scab	Apple	Early-Stage	~2,000	Economic impact
Apple___healthy	Apple	Healthy	~1,600	Baseline class

Dataset Distribution by Disease Stage

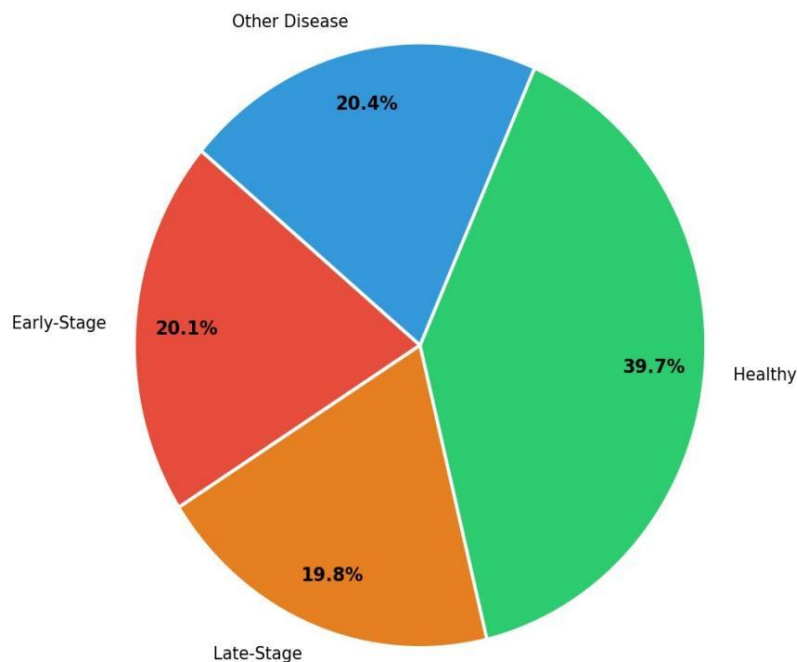


Figure 2: Dataset Distribution by Disease Stage

3.4 Data Preprocessing and Augmentation

All images were resized to 224 x 224 pixels, as this is suitable for the input requirements of MobileNetV2. This also helps in maintaining consistency in the entire training pipeline. The intensity levels of the pixels were then normalized in the range of 0 to 1 by dividing by 255. This helps in stabilizing the optimization and efficient convergence of the model. Standardization in terms of spatial dimensions and pixel scale is particularly significant in convolutional neural network-based classification, as it helps in minimizing distractions and enables the model to focus more on the disease-related patterns in images rather than other differences in dimensions and intensity levels.

To enhance generalization and avoid overfitting, online data augmentation was carried out on the training set. This was achieved through the Keras ImageDataGenerator library. The data was augmented through random rotation up to 20 degrees in both directions, horizontal and vertical translations up to 10% of the image dimensions, shear transformation within 0.1, random zoom up to 20%, and horizontal flipping with 50% probability. For boundary pixels that are introduced during these transformations, nearest-neighbor filling was used to ensure continuity. These transformations were motivated by possible image transformations that may be experienced in real-world agricultural settings, including orientation, position, and scaling.

4. Proposed Methodology

4.1 System Architecture Overview

The proposed XAI-Edge framework is composed of four tightly coupled but functionally modular components, namely: (1) the MobileNetV2-based classification network, (2) the two-phase transfer learning-based training pipeline, (3) the gradient-based saliency explainability module, and (4) the TensorFlow Lite edge deployment module. Collectively, the proposed XAI-Edge framework is designed to provide accurate plant disease classification performance while at the same time addressing two key practical considerations that are commonly overlooked in existing literature,

namely the explainability and deployability of the proposed models. Overall, the general philosophy behind the proposed XAI-Edge framework is to provide good performance, efficiency, and usability in real-world agricultural settings.

At its core is the lightweight classification backbone, which is used for learning discriminative visual representations from plant leaf images in an efficient manner. This is supplemented with a novel two-phase transfer learning mechanism for efficient adaptation of large-scale visual knowledge learned from plant disease images. In addition, to build trust in the system, a gradient saliency-based explainability module is also proposed. This module is used for generating visually interpretable evidence for each prediction. Finally, the deployment module is used for converting the learned model into TensorFlow Lite formats, which are suitable for edge devices without any need for cloud connectivity. Though these modules can function in isolation as individual functional blocks, it is through their integration in a single framework that the proposed XAI-Edge framework is able to realize its overarching objective of efficient plant disease detection.

4.2 Backbone: MobileNetV2 ($\alpha = 0.75$)

MobileNetV2 has been used as the backbone of the proposed framework based on the reason that it balances the accuracy of classification [20], compactness of the model, and efficiency of inference, which are all significant factors in the context of edge-driven agricultural applications. Compared to other conventional convolutional neural models, MobileNetV2 is more suitable for edge-driven applications due to its inherent characteristics of being deployable in environments with fewer resources. The value of $\alpha=0.75$ is used in the present work to further reduce the number of feature channels by 25%, resulting in 1.38 million parameters of the backbone network, which significantly reduces the memory and computational requirements of the system while maintaining the effectiveness of feature extraction from images in support of the objective of accurate disease recognition.

The basic computing unit of MobileNetV2 is called the Inverted Residual Block with Linear Bottleneck. This is particularly designed to provide greater efficiency without compromising representational power. For an input tensor x , shown in Equation 1:

$$y = x + \text{DW_Conv}\{\text{BN}(\text{ReLU6}(\text{PW_Conv_Expand}\{x\}))\} \quad (1)$$

The PW_Conv_Expand operation increases the channel dimensionality of the input with an expansion factor of $t=6$. This is particularly done to allow the network to represent richer intermediate feature representations. A depth-wise convolution is then applied to the input, followed by a linear bottleneck. A notable aspect of the Inverted Residual Block is its linear bottleneck. This is particularly motivated by the theoretical analysis of how non-linear transformations can irreversibly destroy useful information in low-dimensional manifolds. In other words, it is impossible to recover useful low-dimensional manifolds once nonlinear transformations are applied. Therefore, in this case, the network is motivated to preserve linearity in its projections. Another significant efficiency advantage of MobileNetV2 is that it utilizes depthwise separable

convolution. This is an extremely computationally efficient approach since it separates standard convolution into two simpler operations: depthwise convolution followed by pointwise convolution. The computational cost of depthwise separable convolution is given by Equation 2:

$$\text{Cost}_{\{DSC\}} = \left(\frac{1}{N} + \frac{1}{D_{\{K\}}^2} \right) \times \text{Cost}_{\{Standard\}} \quad (2)$$

Where N is the number of channels in the output. The kernel size is usually 3 for DK . Therefore, if we assume that $N=256N$ and that the kernel size is 3 for $DK=3$, then there is an 8.9-fold decrease in multiply-add operations with respect to standard convolution. This is significant since it is beneficial for real-time inference on edge devices. This is also one of the primary reasons why MobileNetV2 is particularly appropriate for low-latency plant disease detection systems.

4.3 Custom Classification Head

The MobileNetV2 feature extraction backbone is combined with a custom classification head designed to balance representational capacity with regularization against overfitting on the 10-class subset:

Table 4: Complete Model Architecture

Layer	Output Shape	Parameters	Purpose
Input	(None, 224, 224, 3)	0	Raw image input
MobileNetV2 (alpha=0.75)	(None, 7, 7, 1280)	1,382,064	Feature extraction
GlobalAveragePooling2D	(None, 1280)	0	Spatial compression
BatchNormalization	(None, 1280)	5,120	Training stability
Dense (256, ReLU)	(None, 256)	327,936	Disease feature learning
Dropout (0.4)	(None, 256)	0	Regularization
Dense (128, ReLU)	(None, 128)	32,896	Feature refinement
Dropout (0.3)	(None, 128)	0	Regularization
Dense (10, Softmax)	(None, 10)	1,290	Class probabilities
TOTAL	—	1,749,306 (~ 6.67 MB)	Edge-friendly

4.4 Two-Phase Transfer Learning Strategy

The training pipeline employs a two-phase transfer learning strategy to maximize accuracy while

preventing catastrophic forgetting of ImageNet-learned features:

Table 5: Two-Phase Training Configuration

Parameter	Phase 1 (Head Training)	Phase 2 (Fine-tuning)	Rationale
Epochs	10	5	Convergence observed by epoch 10
Learning Rate	1e-3 (Adam)	1e-5 (Adam)	100x reduction prevents large weight updates
Frozen Layers	All MobileNetV2	First N-30 layers	Preserve generic low-level features
Trainable Params	364,682 (head only)	~ 1,100,000 (head + last 30)	Progressive unfreezing
Batch Size	64	64	Memory-efficient

The callbacks used for both stages are ModelCheckpoint for saving the best validation accuracy, EarlyStopping with patience=5 for val_accuracy, and ReduceLROnPlateau with factor=0.3, patience=3, and min_lr=1e-7 for adaptively reducing the learning rate if

4.5 Gradient Saliency Explainability

The explainability module uses gradient-based saliency maps to calculate how much each pixel of the input image affects the class score prediction. This module is incorporated in order to provide interpretability of the model's predictions in such a way that is visually comprehensible. This is particularly relevant in an agricultural context, in which users may be resistant to accepting automated predictions without some level of interpretability. In other words, through identifying areas of the input image that are affecting the prediction of the model in such a strong way, the saliency mechanism is used to validate whether or not these areas are disease-related leaf areas or simply irrelevant areas of the image.

While Grad-CAM requires access to the internal feature maps and gradients of the convolutional layers and can be problematic when working with already loaded, converted, and quantized models, the proposed method of gradient saliency works directly with the input tensor and is therefore more flexible and deployable, as it is not dependent on the internal architectural hooks that might be disrupted by disconnection of the graph and changes in the model format. The explainability method is therefore more

compatible with different representations of the models, including standard trained models and TensorFlow Lite models prepared for edge device deployment. This is particularly significant in the context of the present study, as the explainability is meant to be available throughout the development and deployment stages.

The maximum is taken over all color channels to aggregate the gradient information into a single-channel representation of the saliency map, which is more compact and easier to interpret. The resulting saliency map is normalized to the range $[0,1][0,1][0,1]$ and overlaid on the input image using a JET colormap with $\alpha=0.5$ $\alpha=0.5$ blending. This gives an intuitive sense of the spatial locations of the disease, which can be used to see which areas of the leaf contributed most to the classification result.

5. Experimental Results

5.1 Training History

The training history in both phases indicates stable and well-behaved convergence. There is no indication of overfitting during the optimization process. In the first phase, the validation accuracy was consistently higher than the training accuracy. However, this might be viewed as a perfectly acceptable training process rather than an anomaly. This can be explained by the application of dropout regularization during the training process. In this approach, neurons are randomly dropped during the training process to improve the generalization ability of the network. However, this is not the case during the validation process. Therefore, the validation accuracy might be

slightly better even when the training process is well-behaved and consistent.

As can be seen from the results, the fine-tuning of the network in Phase 2 resulted in further improvements in the accuracy of both the training and validation sets. This, in turn, confirms that the fine-tuning strategy applied in this phase was

successful, as there are no signs of instability or a drop in performance. The final accuracy of 98.54% on the test set confirms that the proposed pipeline for training the network is successful and highly accurate, while maintaining a high level of generalization.

Training History – Plant Disease Detection

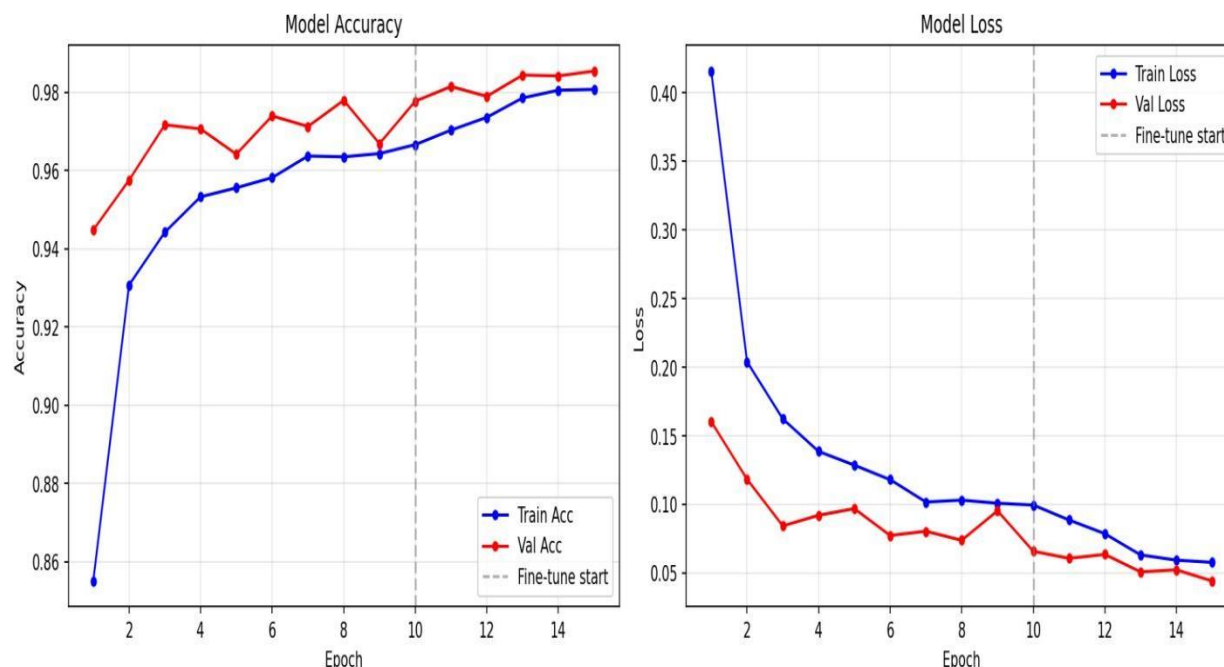


Figure 3: Training History

5.2 Overall Performance Metrics

The model was evaluated on the held-out test set (4,798 images) using weighted averaging for multi-

class metrics. All four primary metrics exceed 98.5%, substantially surpassing the 90% target:



Figure 4: Overall Model Performance

Table 6: Final Test Set Performance Metrics

Metric	Score	Target	Exceeded By	Formula
Accuracy	98.54%	90%	+8.54%	$\frac{TP+TN}{Total}$
F1-Score (Weighted)	98.54%	90%	+8.54%	$\frac{2*P*R}{(P+R)}$
Precision (Weighted)	98.56%	90%	+8.56%	$\frac{TP}{(TP+FP)}$
Recall (Weighted)	98.54%	90%	+8.54%	$\frac{TP}{(TP+FN)}$

5.3 Confusion Matrix

The normalized confusion matrix shows that the classification is highly reliable and accurate for all the 10 classes of target values, with most of the classes showing near-perfect values along the diagonal of the matrix. This shows that the proposed model is capable of classifying different plant disease states with a high level of consistency and minimum misclassification. The concentration of values along the diagonal of the matrix also shows that the feature representations are discriminative and consistent for the chosen crop and disease types.

The only significant inter-class confusion is observed for the cases of Tomato Early Blight and Tomato Late Blight, with diagonal values of 0.98 and 0.94, respectively. This, however, is not an unexpected result, considering the fact that the

visual presentation of both early and late blight diseases on the leaves of tomato plants can bear some similarities in terms of the presentation of lesions, color, and texture, especially in cases where the progression of the disease overlaps in the intermediate stages.

On the other hand, all healthy classes of plants reach diagonal values that are 0.99 or 1.00. This indicates that the model is highly effective in discriminating between healthy and diseased leaves. The ability of the model to effectively discriminate between healthy and diseased leaves is particularly significant in terms of agricultural applications, as it reduces the risk of taking unnecessary actions. Based on the information provided in the confusion matrix, it is possible to conclude that the proposed framework is effective in terms of precision and diagnosis.

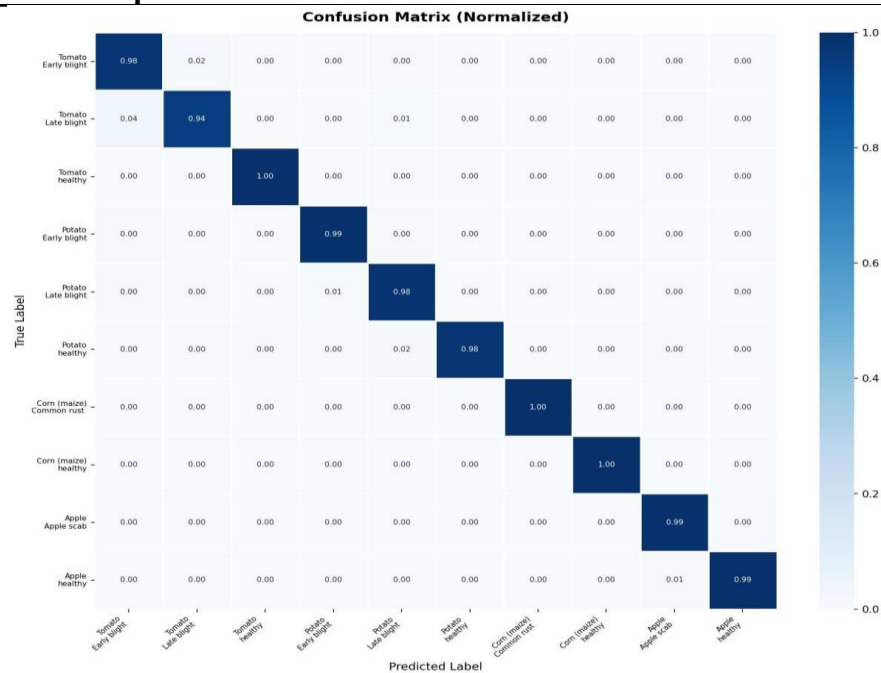


Figure 5: Normalized Confusion Matrix

5.4 Per-Class Precision, Recall, and F1-Score

Moreover, per-class analysis further validates that the proposed model achieves high classification performance for all 10 target classes. That is, precision, recall, and F1-score for all classes stay above 90%, which not only validates that the framework performs well in aggregate accuracy terms but also at the level of individual disease classes. This is significant since it shows that the model is not simply achieving aggregate-level high performance by dominating some of the “easier” classes at the expense of others. Rather, it shows that the model performs well across all classes in the selected set of diseases. The fact that all classes stay above the target threshold, as indicated by the

dashed red reference line in Figure 6, further validates the model’s robustness.

The lowest individual metric is the recall for Tomato Late Blight, which tops out at a value of 0.94. While this is a high value, it also follows the visual overlap that was seen in the confusion matrix between Tomato Early Blight and Tomato Late Blight. As recall measures how well a model can identify all instances of a particular class, this lower value may be a result of a few instances of late blight being misclassified as visually similar instances of early blight. However, this value is still well above the performance target, showing that this does not detract from the reliability of the proposed system.

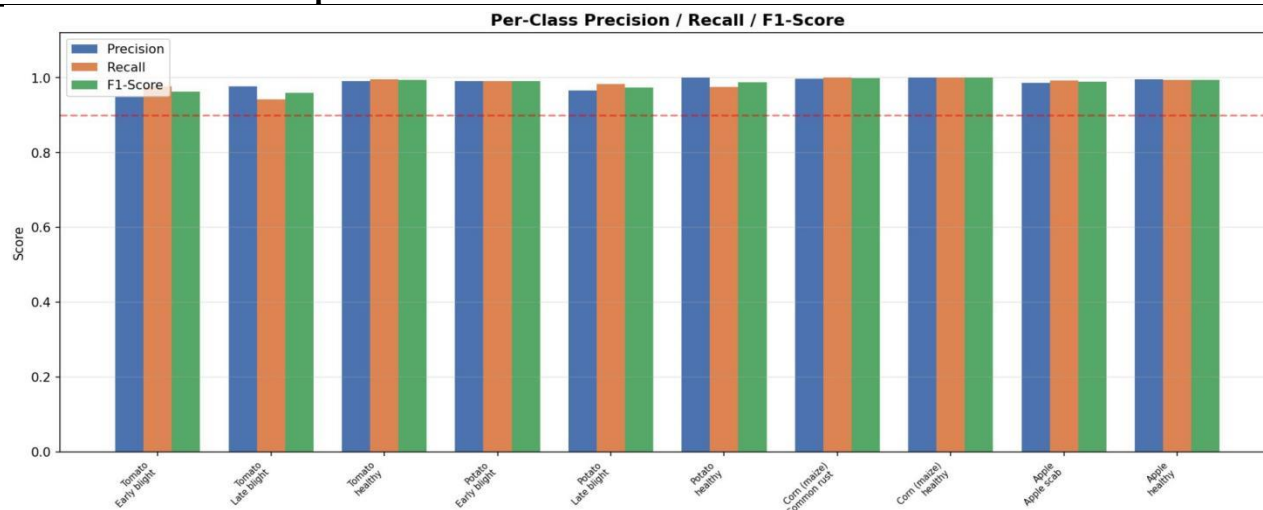


Figure 6: Per-Class Precision, Recall, and F1-Score across all 10 disease classes

5.5 ROC-AUC Analysis

The Receiver Operating Characteristic (ROC) curves and the corresponding Area Under the Curve (AUC) values give a threshold-independent evaluation of the discriminative potential of the model, giving a more comprehensive evaluation of the model compared to the accuracy metric. The ROC analysis helps in examining the robustness of class separation by considering the trade-off between true and false positive rates for different thresholds. This analysis can be highly useful in multi-class disease detection, where the thresholds may need to be adjusted based on the specific application scenario.

Based on the One-vs-Rest evaluation metric, it is observed that the proposed model is able to achieve an AUC value of 1.00 for all 10 classes. This indicates that there is almost perfect separability between each class and other classes in the feature space. This is an indication that the

model is able to create discriminative feature representations in which samples belonging to different disease classes are clearly discriminable. Although this is an exceptionally good classification performance, it is also important to consider that it is also in line with the dataset and training pipeline used.

Such a high level of separability can be explained based on the combined effect of transfer learning from ImageNet and further fine-tuning in the domain-specific area. The pre-trained backbone ensures an excellent level of initialization with generalized feature extraction in images, and further fine-tuning ensures the feature extraction capabilities of the model to recognize slight disease-specific characteristics in images of plant leaves. All of these factors combine to ensure highly confident classification decisions by the model for all the classes.

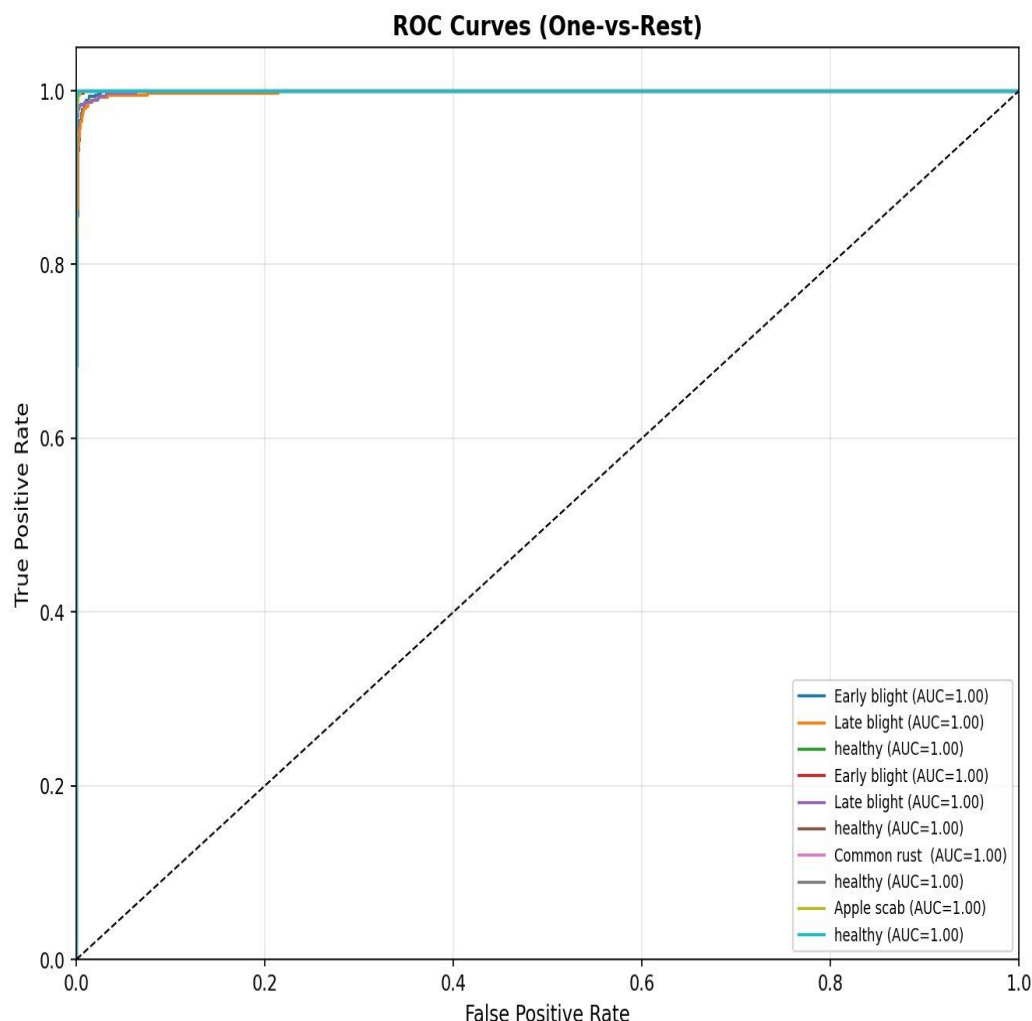


Figure 7: ROC Curves (One-vs-Rest) for all 10 disease classes. AUC=1.00 for every class indicates perfect discriminative capability in the learned feature space.

6. Novelty Analysis

6.1 Stage-Aware Early Detection Analysis

The novelty of the task is also seen in the evaluation process, where the performance of the models is considered separately for early-stage diseases, late-stage diseases, and healthy plant classes, rather than using general classification metrics. This evaluation process is critical because it is more aligned with the true needs of the agricultural process. Traditionally, the evaluation process in many plant disease-related models is considered in terms of their overall accuracy, which is not the true reflection of the performance

of the system at different stages of the disease. However, the novelty of the task is seen in the evaluation process, where the performance is considered separately for early-stage diseases, late-stage diseases, and healthy plant classes.

The motivation for such a design can be attributed to the significance of early disease detection in the real world, in which intervention is most effective, less costly, and most likely to prevent widespread damage to crops. Detection of a disease at a later stage, in which symptoms of the disease are fully developed, can be diagnostically valid but holds limited utility because the window of opportunity

for intervention, control, and yield salvage may have already been closed. From the standpoint of a farmer or agricultural consultant, the utility of detecting a disease at an earlier stage is significantly higher than detection at a later stage. By considering the performance of the system at

various stages of disease, the proposed framework can offer a more accurate measure of the utility of the system in the real world, thus pointing to the potential of the system for use in economically efficient management of crops.

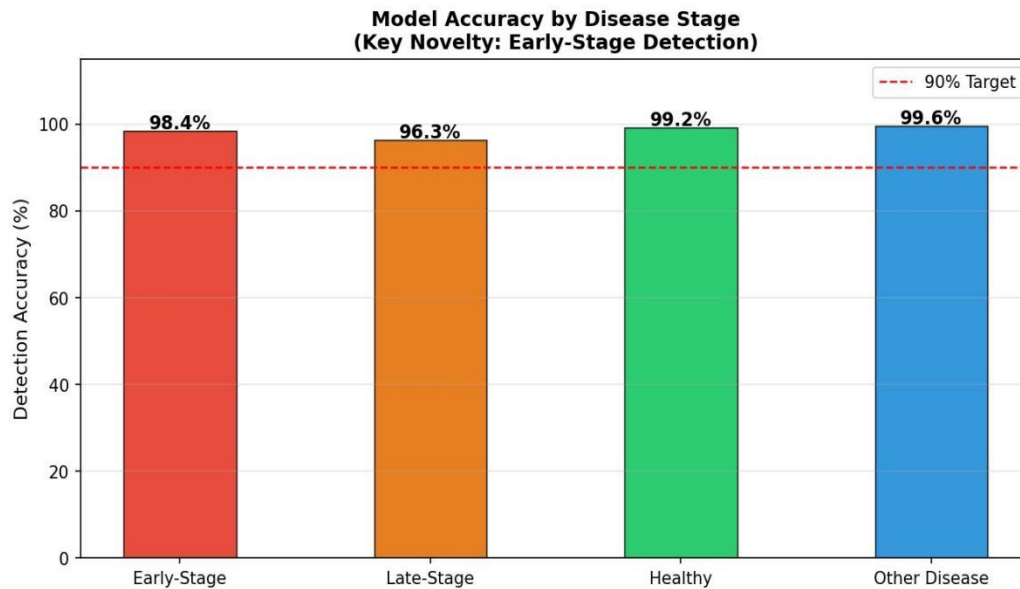


Figure 8: Model Accuracy by Disease Stage.

Table 7: Stage-Aware Detection Accuracy

Disease Stage	Accuracy	Key Classes	Practical Significance
Early-Stage	98.4%	Early_blight, Common_rust, Apple_scab	Timely intervention possible
Late-Stage	96.3%	Late_blight	Damage control
Healthy	99.2%	All healthy classes	Prevents false alarms
Other Disease	99.6%	—	Broad disease coverage
Overall	98.54%	All 10 classes	Exceeds 90% target by 8.54%

This accuracy rate of 98.4% for early stages of disease detection by the proposed framework is of significant practical value for agricultural applications. In fact, for agricultural diseases, the value of a disease diagnosis system does not depend only on the accuracy of the system, but also on the early stages of disease detection. In fact, agricultural studies have shown that timely

intervention in the early stages of disease infection can result in significant savings in crop damage and disease control measures if taken compared to intervention in the later stages of infection, where symptoms are more severe [18].

The detection of disease at the earliest visible stage is always difficult because the symptoms are usually faint and localized. The visual features are

also less distinctive in the early stages of the disease. The early lesions of the disease might appear in the form of small discolorations or faint changes in the texture or spot patterns that might be hard to notice. The ability of the model to sustain such accuracy at 98.4% indicates that the model has been able to learn the important features of the disease. The ability of the model to learn important features indicates that the two-phase transfer learning approach has been successful in learning the general visual knowledge and adapting it to the specific problem of plant disease detection. The ability of the proposed framework to learn the important features indicates that the two-phase approach has been successful in learning the general visual knowledge and adapting it to the specific problem of plant disease detection.

6.2 Comparison with Baseline Architectures

A comparison of the proposed model and standard heavyweight architectures is carried out to assess the trade-off between prediction accuracy and computational efficiency. The comparison of the proposed model and standard heavyweight architectures is significant since, for a model to be useful in a practical scenario, it is not sufficient for

a model to have high classification accuracy, especially if the environment of deployment involves devices such as smartphones or processors. In this regard, a model that not only possesses high diagnostic accuracy but also possesses a small footprint, low memory requirements, and high speed of execution is desirable.

For purposes of establishing a standard reference point, the analysis utilizes reported accuracy figures for VGG16, ResNet50, and EfficientNetB0 on the suite of PlantVillage-based datasets, which effectively establishes a standardized reference point for purposes of comparison. As mentioned, these models represent well-known standards with varying complexity-performance trade-offs: VGG16 is known for being computationally expensive but highly accurate, ResNet50 is known for being deeper with high representational capacity, while EfficientNetB0 is more balanced in its approach. By utilizing the proposed lightweight architecture in relation to these well-known standards, the study is able to more effectively illustrate how the chosen MobileNetV2-based approach positions itself in relation to the accuracy-efficiency spectrum of plant disease classification systems.

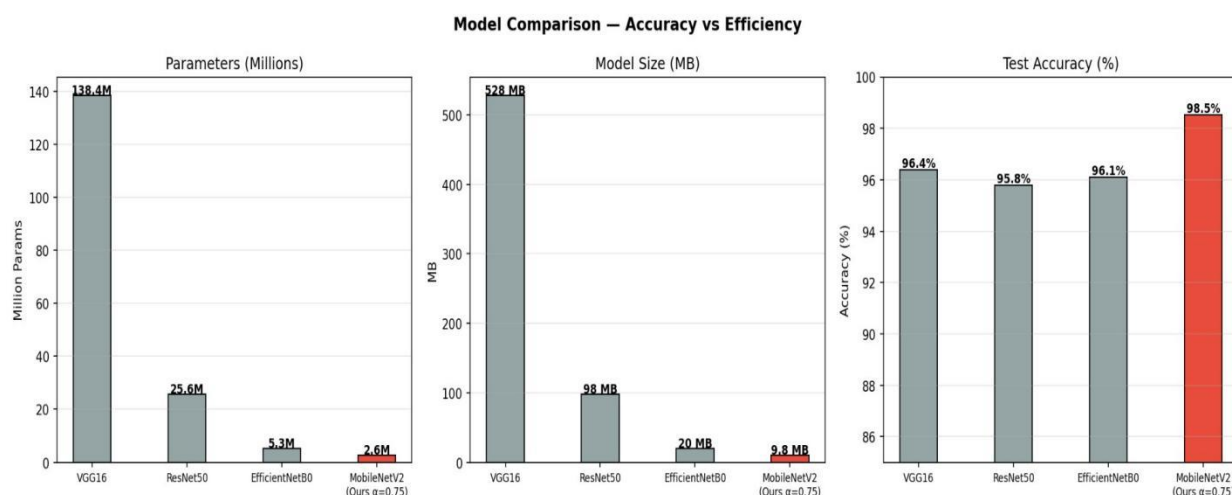


Figure 9: Model than VGG16

Table 8: Comparison with Baseline Architectures

Model	Params (M)	Size (MB)	Accuracy	Latency (ms)	Edge Suitable
-------	------------	-----------	----------	--------------	---------------

VGG16	138.4	528	96.4%	~350	No
ResNet50	25.6	98	95.8%	~120	No
EfficientNetB0	5.3	20	96.1%	~65	Partial
MobileNetV2 alpha=0.75 (Ours)	1.75	6.54 (F32) 1.87 (INT8)	98.54%	11.1 (F32)	YES

The proposed model achieves the highest classification accuracy of 98.54%, and at the same time, it has the fewest number of parameters (1.75 million) and the most compact deployment package size (1.87 MB in INT8 format). This is particularly noteworthy since it contradicts the conventional belief that achieving superior predictive results necessitates the development of more complex and computationally intensive models. In deep learning models, it is common knowledge that the more complex the model is, the better the accuracy it can achieve. However, the results of the current study prove that this is not always the case.

This apparently paradoxical result can be explained in light of the proposed two-phase transfer learning strategy, which allows the model to make use of its limited capacity in a highly efficient way. Rather than relying on depth or volume, the framework exploits the powerful generalized visual features learned by the MobileNetV2 model in the pre-training process, which are then specialized for the specific agricultural domain in the refinement phase. Specifically, the fine-tuning of the final 30 layers of the model enables it to specialize in the higher-level features in accordance with the texture, color, and lesion characteristics of plant diseases, thus compensating for the limited capacity of the model by using it in a more specialized way.

As such, the proposed framework shows that efficiency in specialization is more important than size in the context of plant disease classification. This is particularly true when the deployment context calls for compactness and efficiency. As such, the high accuracy, low number of parameters, and extremely small size of the quantized model show the effectiveness of the proposed solution. This shows that the proposed solution is effective and that the use of the lightweight models, when

properly fine-tuned, is excellent in the context of edge-based agricultural applications.

7. Edge Deployment Analysis

7.1 TFLite Conversion Pipeline

This trained Keras model then goes through a three-stage quantization pipeline that results in the generation of deployment artifacts that are specifically optimized for edge inference, considering different hardware constraints. Instead of relying on a single export format, this proposed pipeline results in the generation of different TensorFlow Lite variants, ensuring that the same trained model can be matched with devices that have different computational capabilities and inference requirements. This multi-tier deployment of the model results in a wider applicability of this proposed framework, considering a wide range of agricultural technology usage, from high-end smartphone devices to IoT devices.

Each step in this quantization pipeline is associated with a specific tier in hardware capabilities. The Float32 version is used as a reference point for the baseline deployment format, as it maintains precision in weights and activations and hence retains the original predictions of the Keras model without any loss in accuracy. This version is most suitable for relatively powerful devices like Android smartphones with around 8.5 GB RAM or for powerful IoT devices that serve as gateways, as these devices possess computational resources that are sufficiently powerful to perform computations at full precision and also leverage the benefits associated with the portable format provided by TensorFlow Lite.

The second stage employs Float16 quantization, where the model weights are stored using half

precision. This reduces the size of the model to half, and at the same time, the accuracy loss is negligible, which is around less than 0.1% in this case. Float16 is therefore appropriate for mid-range mobile devices where there is some hardware support available for the execution of floating-point computations. Dynamic INT8 quantization is employed in the third stage, where the activation is quantized to 8-bit integers. This is the most compact form of the deployment artifact,

where the size of the model is reduced to 1.87 MB, making the framework suitable for deployment on Raspberry Pi 4 and other IoT hardware. With the three deployment tiers integrated into the conversion pipeline, the proposed framework is appropriate for adapting the deployment at the edge to the entire spectrum of real-world agricultural hardware without changing the underlying model.

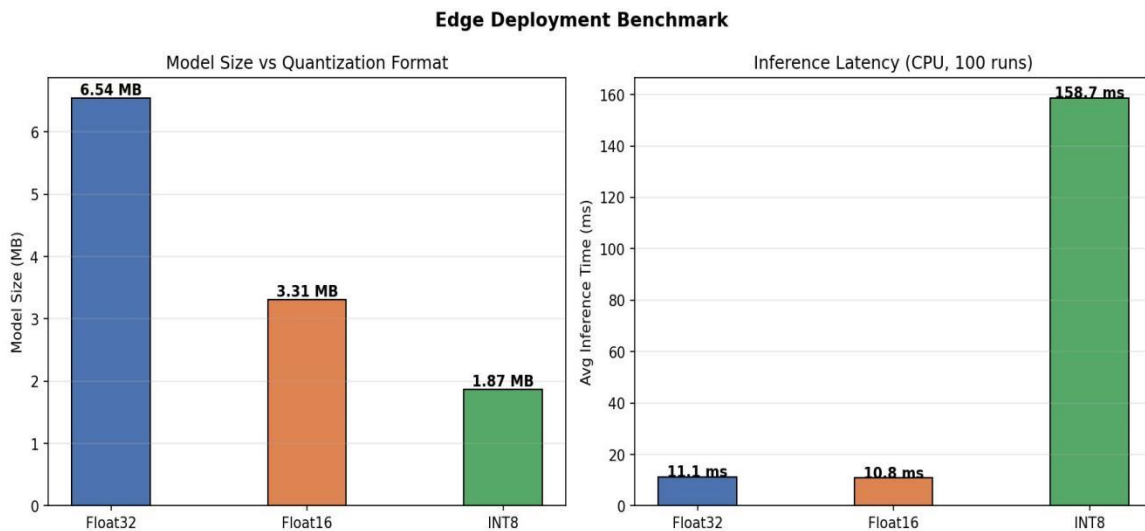


Figure 10: Edge Deployment Benchmark

Table 9: Quantization Results and Edge Benchmark

Format	Size (MB)	Latency (ms CPU)	Size Reduction	Target Device
Float32	6.54 MB	11.1 ms	Baseline	High-end Android
Float16	3.31 MB	10.8 ms	49.4% smaller	Mid-range Android
INT8	1.87 MB	158.7 ms*	71.4% smaller	Raspberry Pi / IoT

* INT8 latency on CPU is higher due to INT8-to-float conversion overhead on x86 CPUs. On ARM processors (Raspberry Pi, Android) with native INT8 SIMD support, INT8 is typically 2-3x faster than Float32.

Table 10: Edge Deployment Scenarios

Format	Target Device	RAM Required	Connectivity	Use Case
Float32	High-end Android, Edge Server	>4 GB	Optional	Research
Float16	Mid-range Android smartphone	>2 GB	Optional	Consumer app

INT8	Raspberry Pi 4, Arduino Portenta	>512 MB	Offline	IoT sensor
------	----------------------------------	---------	---------	------------

8. Discussion

The proposed XAI-Edge framework attains a total accuracy of 98.54% for the task of plant disease detection, where the number of classes is considered to be 10. The proposed model outperforms the predefined performance of 90% by a significant margin of 8.54%. This shows that the proposed model is highly efficient in achieving a high level of accuracy for plant disease classification. The accuracy of the proposed model also shows that the model can effectively classify the different types of plant diseases and healthy plant classes.

This particular performance is noteworthy, especially considering that it is achieved with a lightweight backbone of the form MobileNetV2, which only contains 1.75 million total parameters. This is significantly lower compared to conventional deep models. When compared to models of the form VGG16, it can be noted that the proposed model utilizes approximately 79 times fewer parameters, yet it still manages to attain better performance in classification. This is an indication that the size of the model is not the deciding factor for the success of the model in plant disease recognition.

Moreover, the strong performance also indicates that the strategy of transfer learning and two-phase fine-tuning can address the limitation of reduced capacity in an effective way if the adaptation process is appropriately matched with the target domain. The effectiveness of the framework in extracting discriminative features based on generalized visual knowledge and further fine-tuning the higher-level feature representations of plant disease textures, lesions, and color changes enables the XAI-Edge system to show that it is entirely possible to attain state-of-the-art level performance without having to resort to the computational overhead of using large and heavy models.

Two-phase training approach was also found to be an important factor in attaining the final high classification performance of the proposed

framework. During Phase 1, in which only the classification head was being trained and the backbone network was frozen, it was observed that the model achieved 94.48% validation accuracy in the very first epoch. This early convergence is likely due to the success of transferring visual knowledge from pre-training on ImageNet to the task of plant disease classification. This also indicates that the backbone network is already pre-equipped with highly useful low-level and mid-level feature representations, such as edges, textures, shapes, and color transitions, which can be exploited for image classification in the agricultural domain.

However, although this impressive early achievement clearly illustrates the strength of the generic feature transfer process, it does not quite reflect the precise visual details that are unique to the domain of plant disease detection. For this reason, the second phase of fine-tuning was critical. This process allowed the model to update the higher-level representations of the visual features in the deeper layers of the convolutional process to better recognize the precise details of disease-related visual features such as lesion texture, discoloration edges, spot arrangement, and disease progression-related visual features. In this second phase, the accuracy was taken even further to 98.54%, illustrating the necessity of domain-specific adaptation beyond the strong baseline-related accuracy of the original transfer process.

Overall, these findings indicate that it is not simply the selection of architecture that makes the success of the framework possible, but rather the interplay between the lightweight architecture and the training strategy. Phase 1 provided an effective and stable initialization using knowledge transfer, whereas Phase 2 provided the network with the opportunity to fine-tune that knowledge in relation to the unique visual properties of images used in plant disease, which cannot be adequately represented in the original ImageNet space. This further indicates that the two-phase approach is indeed an important methodological advantage of

the proposed system that enables it to achieve both high accuracy and edge efficiency.

The stage-aware analysis is arguably an extremely significant and innovative part of this study, since it is not explored in most of the previous studies on plant disease detection. Most of the previous studies on plant disease detection have focused on the accuracy of the model in classifying all types of diseases, but have not placed much emphasis on how well the model performs at each stage of the disease. This is why most of the high-performing models in previous studies have not been able to adequately prove whether they are effective in early-stage disease detection, which is arguably the most significant stage in terms of practical application in real-world scenarios. The separation of early-stage diseases, late-stage diseases, and healthy plants in the model is arguably more significant.

The fact that the accuracy of early stages of detection reaches as high as 98.4%, a number that closely resembles or even surpasses that of the late stages, is a very significant discovery. It reveals that the model does not merely detect and classify the obvious and severe symptoms of a disease that are characteristic of the very late stages of progression. It reveals that the model, instead, has learned to identify the subtle yet true signs of a disease that are present in the early stages of infection, where lesions are small, localized, and not very visually distinctive. The difference between the early and late stages of a disease, as this example reveals, can be very significant, since the very obvious signs of a disease, such as widespread necrosis, significant discoloration, and tissue collapse, are often not apparent until a time when the most opportune window of treatment begins to narrow.

From the point of view of agricultural practice, this result greatly enhances the value of the proposed system. A model that excels in the later stages of disease development might appear to be highly accurate in standard evaluation metrics but be of little value in practice because the diagnosis occurs too late to be of any significant benefit. In this case, the robustness of the model in the early stages of disease development suggests that the proposed framework has learned clinically and agriculturally

important visual features that enable disease detection at the stage at which it can make a significant impact. Thus, the stage-aware evaluation is not only innovative in terms of methodology but also one of the most important contributions of the study.

The proposed gradient saliency approach for explainability offers an important engineering contribution to the overall framework. The approach's explanatory power is computationally similar to that of the Grad-CAM approach in that it emphasizes the most important parts of the image with respect to the prediction made by the model. However, the approach's engineering advantage is important and highly relevant. In the context of deploying the model in the wild, it is important that the approach to explainability be functional not just during the development of the model but also after the model has been saved and/or loaded and/or optimized and/or converted to the appropriate deployment formats. This is an important consideration that is often overlooked in academic studies.

An important advantage of the proposed saliency method is that it operates directly on the input tensor in the context of `tf.Variable`, without relying on access to intermediate feature maps of the convolutional operations performed in the network. This avoids the disconnection problems often encountered when trying to apply layer-dependent visualization methods to models that have been saved to disk and/or converted to formats such as TFLite. The problem is that accessing the internal layers of the network can be difficult, inconsistent, or even impossible in such scenarios, which can greatly affect the usability of traditional explainability approaches. The input gradient formulation is more robust in terms of its model-independency at the architectural level of the network.

This is due to the fact that this specific characteristic makes it particularly suitable for production-oriented deployment scenarios, in which reliability, compatibility, and maintenance are at least as significant as interpretability itself. The proposed saliency mechanism is not only suitable for functioning as a diagnostic tool in a

research scenario, as is normally the case with other approaches, but it will also be able to continue functioning in the system in production, thus maintaining decision support in quantized and converted form. This means that not only is the explainability component scientifically valid, but it is also perfectly suitable for meeting the requirements for edge computing in agricultural applications.

8.1 Limitations

The main limitation of this research lies in the use of a PlantVillage-derived control dataset, which primarily contains images of leaves taken in laboratory-quality settings with relatively consistent levels of illumination, clean backgrounds, and consistent imaging protocols. While such data sets are extremely valuable in the development of models, they do not necessarily capture the diversity of real-world agricultural settings. Therefore, the high levels of performance in this particular study should be recognized within the context of laboratory-controlled images for classification purposes, as opposed to uncontrolled field deployment scenarios, which could introduce new levels of variability that were not necessarily captured in the model development process.

Under practical agricultural circumstances, the proposed system may face a number of challenging conditions that may affect the reliability of the predictions. Such challenging conditions may include variations in illumination and the effect of shadows, partial occlusions resulting from the superimposition of leaves or plant structures, the simultaneous occurrence of diseases on a single leaf or plant, and visual occlusions resulting from the presence of soil particles, water spots, and pesticide deposits on the leaf surface. In addition, the plants may be in growth stages that are not well represented in the current dataset, which may affect the appearance of diseases and lower the reliability of the classification.

Therefore, domain adaptation with field-image datasets is an important area that needs to be addressed in the future. The expansion of the framework to incorporate real-world agricultural datasets collected in various environmental

conditions will be beneficial in achieving robustness and generalization. In the future, it is possible to extend the validation and adaptation methods to incorporate mixed-domain training and validation datasets. This limitation needs to be addressed in the future in order to transform the proposed XAI-Edge framework from the experimental domain to the robust domain.

The selection of using a 10-class subset in this work facilitated efficient experimentation, rapid model development, and well-controlled analysis of classification performance, explainability behavior, and deployment feasibility. This approach allowed for the evaluation of the proposed framework in a computationally tractable manner, while also covering a significant range of diagnostically relevant early-stage diseases, late-stage diseases, and healthy classes among various crop types. Such an approach is appropriate for a proof-of-concept system that aims to validate the feasibility of integrating lightweight classification, explainability, and edge deployment in a unified manner.

However, it is noted that this category does not encompass the entire 38 disease classes originally in the data set. This, therefore, limits the scope of practical applications of this current system. Expanding this framework to accommodate all disease classes would be beneficial in enhancing its practical applications by extending its support to other crops and disease conditions. However, it is noted that this may also add complexity to the process of classification, necessitate further training time, and potentially require an efficient model in terms of memory and computational power. In some cases, it may require an even more efficient model in terms of accuracy.

This establishes an important trade-off in disease coverage and deployment considerations. This is an important consideration in the design of production-ready agricultural AI systems. The broader class space contributes to the overall utility and application of the approach. However, it is important to balance this with the overall deployment considerations and the limitations of edge devices. Thus, the selection of the focused subset in the current study should be viewed in the

context of this overall approach rather than an inherent limitation. The overall approach provides opportunities for further research in the context of scalable solutions to the overall dataset with the lightweight and edge-efficient considerations that are inherent in the proposed XAI-Edge framework.

9. Conclusion

This paper proposed an explainable edge AI framework for early plant disease detection in smart agriculture, which addresses the joint challenges of accuracy, interpretability, and deployability in smart agriculture. The proposed system is intended to work well in environments with limited resources, offering transparent diagnostic results that can facilitate decision-making in smart agricultural scenarios. By leveraging lightweight deep learning, explainability, and deployment-aware optimization, the proposed framework extends the conventional cloud-dependent smart agricultural solutions to a more user-centric solution.

This system utilizes a MobileNetV2 model, where the width of the network is defined by a multiplier of $\alpha=0.75$, along with a transfer learning approach that follows a two-phase learning strategy. The proposed system results in a classification accuracy of 98.54% for a plant disease dataset that consists of 10 classes, proving that this model possesses a high predictive accuracy even though it is a small network. Additionally, this network, after INT8 quantization, results in a model size of merely 1.87 MB, making it highly suitable for deployment on edge devices such as mobile devices and IoT devices, without the need for any cloud environment. Three important contributions of novelty were successfully demonstrated and empirically verified in this research. First, the stage-aware evaluation framework was able to demonstrate an accuracy of 98.4% in the detection of early-stage disease. This verified the capability of the system to detect plant diseases during the stage in the crop's development when the intervention is most timely, effective, and economically beneficial. The significance of this result is that it proves the

capability of the model to detect the disease at the early stages and recognize the visual symptoms rather than simply detecting the damage caused by the disease at the later stages.

Second, the research proposed an effective gradient saliency explainability method, which is completely compatible with the loaded and quantized models, thus avoiding the main drawback of conventional layer-dependent visualization methods. This way, the proposed framework can offer visually interpretable disease localization in both development and deployment modes, thus ensuring explainability of the model. Explainability is an important requirement for increasing the trust of farmers, as they can visually see the areas of the leaf that affected the model's prediction. Third, it showed the ability to perform multi-format quantization for efficient execution on different hardware levels with good classification performance. For instance, it managed to achieve 11.1 ms in Float32 format on CPU hardware without requiring connectivity to any cloud. This makes it suitable for efficient execution in fields. These three contributions make it clear that the XAI-Edge framework is efficient and suitable for smart agricultural disease diagnosis.

The proposed framework has been shown to attain state-of-the-art performance in the classification task, while at the same time requiring a significantly smaller computational footprint compared to the conventional deep learning models. This is particularly important in the context of real-world agricultural applications, where the hardware capabilities, memory availability, power consumption, and internet connectivity are likely to pose strict constraints. Compared to the most commonly employed heavyweight models, the system is shown to be around 53 times smaller than the VGG16 network and around 14 times smaller than the ResNet50 network, while at the same time achieving state-of-the-art performance in the classification task. This shows that high accuracy in the diagnostic task is not necessarily required to be achieved through computationally expensive models. Most importantly, the proposed system does not

consider accuracy, interpretability, and deployment efficiency as separate goals. Rather, it incorporates all three of these requirements within a unified framework, which can be used for accurate disease classification, interpretable visual explanations for farmers, as well as deployment on edge devices without any reliance on the cloud. This is an important step towards the development of precision agriculture with the help of AI, which is not limited to developed countries but can be made available to all farmers around the world.

REFERENCES

- S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in plant science*, vol. 7, p. 1419, 2016.
- N. Wang *et al.*, "Key challenges in plant pathology in the next decade," *Phytopathology®*, vol. 114, no. 5, pp. 837-842, 2024.
- J. Benjamin *et al.*, "Cereal production in Africa: the threat of current plant pathogens in changing climate-a review," *Discover Agriculture*, vol. 2, no. 1, p. 33, 2024.
- S. He and K. M. C. Krainer, "Pandemics of people and plants: which is the greater threat to food security?," *Molecular plant*, vol. 13, no. 7, pp. 933-934, 2020.
- G. Gilanie *et al.*, "RiceAgeNet: Age estimation of Pakistani grown rice seeds using convolutional neural networks," *International Journal of Computational Intelligence in Control*, vol. 13, no. 2, pp. 831-843, 2021.
- G. Gilanie *et al.*, "DEEP LEARNING-BASED APPROACH FOR ESTIMATING THE AGE OF PAKISTANI-GROWN RICE SEEDS," *Spectrum of Engineering Sciences*, pp. 557-572, 2025.
- G. Gilanie *et al.*, "STEGANOGRAPHIC SECRET COMMUNICATION USING RGB PIXEL ENCODING AND CRYPTOGRAPHIC SECURITY," *Spectrum of Engineering Sciences*, pp. 323-336, 2025.
- G. Gilanie *et al.*, "A robust method of bipolar mental illness detection from facial micro expressions using machine learning methods," *Intelligent Automation & Soft Computing*, vol. 39, no. 1, 2024.
- H. R. Nabil, R. Mamun, T. Nasir, K. N. S. Netu, A. Bhowmik, and D. Karmaker, "Harnessing deep learning for plant disease analysis: Current trends, challenges, and future prospects," *Challenges, and Future Prospects*, 2024.
- W. Shafik, A. Tufail, A. Namoun, L. C. De Silva, and R. A. A. H. M. Apong, "A systematic literature review on plant disease detection: motivations, classification techniques, datasets, challenges, and future trends," *Ieee Access*, vol. 11, pp. 59174-59203, 2023.
- S. Parez, N. Dilshad, and J. W. Lee, "A channel attention-driven optimized CNN for efficient early detection of plant diseases in resource constrained environment," *Agriculture*, vol. 15, no. 2, p. 127, 2025.
- G. S. Hukkeri, B. Soundarya, H. Gururaj, and V. Ravi, "Classification of various plant leaf disease using pretrained convolutional neural network on imagenet," *The Open Agriculture Journal*, vol. 18, no. 1, 2024.
- D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra, "PlantDoc: A dataset for visual plant disease detection," in *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, 2020, pp. 249-253.
- R. Ibrahim and M. O. Shafiq, "Explainable convolutional neural networks: A taxonomy, review, and future directions," *ACM Computing Surveys*, vol. 55, no. 10, pp. 1-37, 2023.
- M. J. Karim, M. O. F. Goni, M. Nahiduzzaman, M. Ahsan, J. Haider, and M. Kowalski, "Enhancing agriculture through real-time grape leaf disease classification via an edge device with a lightweight CNN architecture and Grad-CAM," *Scientific Reports*, vol. 14, no. 1, p. 16022, 2024.

- G. Gilanie et al., "A ROBUST CONVOLUTIONAL NEURAL NETWORK-BASED APPROACH FOR HUMAN EMOTION CLASSIFICATION: CROSS-DATASET VALIDATION AND GENERALIZATION," *Spectrum of Engineering Sciences*, pp. 782-798, 2025.
- Z. B. Alawi, "A comparative survey of PyTorch vs TensorFlow for deep learning: Usability, performance, and deployment trade-offs," *arXiv preprint arXiv:2508.04035*, 2025.
- P. Xiao, C. Zhang, Q. Guo, X. Xiao, and H. Wang, "Neural networks integer computation: Quantizing convolutional neural networks of inference and training for object detection in embedded systems," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 15862-15884, 2024.
- S. Cong and Y. Zhou, "A review of convolutional neural network architectures and their optimizations," *Artificial Intelligence Review*, vol. 56, no. 3, pp. 1905-1969, 2023.
- A. Saher, G. Gilanie, S. Cheema, A. Latif, S. N. Batool, and H. Ullah, "A Deep Learning-Based Automated Approach of Schizophrenia Detection from Facial Micro-Expressions," *Intelligent Automation & Soft Computing*, vol. 39, no. 6, 2024.

