Scientific Reports

https://doi.org/10.1038/s41598-026-45947-7

Article in Press

Optimized Lightweight U-Net and YOLACT framework for multi-disease severity detection in pome fruit leaves

Received: 13 November 2025

Accepted: 23 March 2026

Published online: 26 March 2026

Cite this article as: Qasim M., Adnan S.M. & Safi Q.G.K. Optimized Lightweight U-Net and YOLACT framework for multi-disease severity detection in pome fruit leaves. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-45947-7

Muhammad Qasim, Syed M. Adnan & Qamas Gul Khan Safi

Optimized Lightweight U-Net and YOLACT Framework for Multi-Disease Severity Detection in Pome Fruit Leaves

Muhammad Qasim [1,2], Syed M. Adnan [1], Qamas Gul Khan Safi [1]

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan; mohammad.qasim@students.uettaxila.edu.pk (M.Q), syed.adnan@uettaxila.edu.pk (M.A), qamas.gul@uettaxila.edu.pk (Q.G)
[2] Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad, Pakistan; muhammad.qasim@szabist-isb.edu.pk (M.Q)
* Correspondence: mohammad.qasim@students.uettaxila.edu.pk

Abstract

The growing global demand for food production, coupled with the increasing threat of plant diseases, necessitates advanced and automated solutions for crop health monitoring. Among various crops, pome fruits such as apples and pears are widely cultivated yet highly susceptible to multiple diseases that can significantly reduce yield and quality. Existing approaches for disease detection and severity classification are often limited by their dependency on manual inspection and their inability to handle complex real-world imagery, especially when multiple diseases coexist on a single leaf.

To address these limitations, this research introduces a novel dual-model deep learning framework for multi-disease severity detection and classification in pome fruit leaves. A fine-tuned MobileNetV2 backbone is employed to extract high-level discriminative features from a specialized pome leaf dataset annotated with multiple disease types and severity levels. The proposed system integrates a lightweight Lite-U-Net for semantic segmentation to isolate diseased regions and an enhanced Lite-YOLACT for instance segmentation using a linear combination of prototype masks and mask coefficients. Moreover, a new multi-disease severity scale is proposed to quantify the impact of multiple coexisting infections on a single leaf, an aspect not addressed in previous studies. To enhance interpretability, an improved Grad-CAM technique generates visual heatmaps highlighting the most influential regions in the model's decision-making process, providing transparency and validation for agricultural experts. Experimental evaluations demonstrate that the proposed framework achieves 95% accuracy in disease severity estimation, effectively identifying and grading multiple infections simultaneously. This study represents a significant step forward in precision agriculture, offering an efficient, interpretable, and scalable deep learning solution for real-world crop health monitoring and management. The source code and trained models are publicly available at: https://github.com/mqasim0787/Multi-Disease-Severity

Keywords: Multi-disease Severity, Disease Severity Classification, Deep Learning, Neural Network, U-Net, Gradient-weighted Class Activation Mapping (Grad-CAM).

1. Introduction

The agricultural industry has increasingly turned to advanced technologies to combat the persistent challenge of plant diseases, which threaten food safety and global agricultural output. Maintaining strong and steady crop yields is becoming more and more important for sustainable subsistence as the world's population grows. Plant diseases, one of the most widespread and economically destructive threats to this crucial industry, are tenacious and formidable [1]. These challenges are met by complex internal signaling; for instance [2], auxins regulate hormonal crosstalk to activate defense mechanisms [3], while reactive oxygen and nitrogen species (ROS/RNS) maintain redox homeostasis during pathogen attacks. Despite these natural defenses, the potential for yield loss remains high, necessitating more accurate techniques to monitor crop health. These diseases have the potential to drastically lower yields, degrade the quality of products, and cause farmers all over the world to suffer large financial losses. Among the wide range of cultivated plants, pome fruits, such as apples and pears, are particularly vulnerable to various diseases that can significantly impact their growth, development, and market value [4]. Historically, agricultural specialists have mostly depended on physical examination for the identification and severity evaluation of these illnesses. This approach is intrinsically restricted by its subjectivity, labor-intensiveness, and vulnerability to human mistake [5]. Timely and effective intervention efforts are hampered by the inability of such traditional methodologies to reliably diagnose illnesses, particularly in the complex and dynamic settings of real-world agricultural ecosystems [6]. This shows the need for a sophisticated and more accurate technique to monitor crop health, which is not very heavy on the system as well, and could be easily implemented in real time environment. These needs demand deep learning to jump in and develop a more accurate, interpretable multi-leaf disease severity classification system. In this regard, we have proposed a novel deep learning technique to effectively detect the multi diseases on a single leaf in pome crops, providing both precision and transparency in disease management.

Our methodology utilizes a dual-model strategy to navigate the complexities of leaf imagery. The framework integrates Lite-U-Net for semantic segmentation and Lite-YOLACT for instance segmentation. This pipeline is motivated by the need to isolate symptomatic regions and handle the visual noise inherent in field-captured images, ensuring the model focuses strictly on relevant pathological features. To improve the clarity of the decision-making process, we employ Enhanced Grad-CAM to generate visual heatmaps. These heatmaps highlight the most influential regions within the

pome leaf images, allowing agronomists to intuitively verify the model's reasoning and make informed treatment decisions. Unlike works that focus only on single-disease classification or detection, the proposed framework targets the underexplored problem of multi-disease severity estimation under co-occurring infections. The novelty lies not in introducing a new backbone network, but in a tightly integrated lightweight dual-model pipeline that couple's segmentation and instance-level analysis with a formally defined multi-disease severity scoring mechanism. The framework is specifically optimized for mixed pome leaf infections, computational efficiency, and deployment feasibility. Our proposed methodology offers a significant advancement over traditional disease detection methods. Through rigorous experimentation and comparative analysis, our approach demonstrates superior performance, achieving high accuracy, recall, and precision metrics in detecting and classifying disease severity in Pome crops.

Our work introduces several key contributions to the field of plant disease detection:

We introduce a novel approach to multi-disease severity classification on single Pome leaves, which addresses the previously unresolved challenge of overlapping diseases, offering a significant advancement in plant pathology. We introduce a unique fusion of Lite-U-Net and Lite-YOLACT architectures, where U-Net is used for precise semantic segmentation to isolate disease-affected regions, and YOLACT is used for segmentation at the instance level and multi-disease classification on a single leaf. This dual-model approach is designed to address the complexities of multiple overlapping diseases on Pome leaves, a challenge that has not been effectively tackled in existing literature. We proposed a novel Multi-Disease severity scale for Pome fruit containing Pear and Apple leaves. This scale is divided into 0-8 levels. 0 is the lowest, 8 is the highest. We incorporate Enhanced Grad-CAM in our methodology to generate interpretable heatmaps, which help in finding areas of focus during disease classification. This approach validates the model predictions and also improves transparency by ensuring that decisions regarding disease management are both accurate and understandable to agronomists and other stakeholders.

These contributions highlight the potential of our approach to transform disease management practices, offering a powerful tool for improving crop health and agricultural productivity.

The structure of the paper is as follows: Section 2 describes the Related work, and the specifics of the suggested methodology are contained in Section 3. Comprehensive experiments along with a detailed comparative analysis with the state-of-the-art frameworks are presented in Section 4. Finally, Section 5 concludes the paper.

2. Related Work

The assessment of plant disease severity is vital for crop management. The traditional methods using handcrafted features have limitations [7]; modern research examines machine learning and deep learning strategies. Studies have enhanced CNNs with global pooling, normalization, and inception modules for enhanced disease severity classification [8], PD2SE-Net, which integrated ResNet50 with shuffle units severity assessment [9, 10]. Using deep learning, a novel method called BR-CNNs is used for agricultural leaf diseases. LP-CNNs and MLP-CNNs were surpassed by BR-CNNs, utilizing CNN techniques within a binary relevance multi-label learning framework [11]. VGG16 networks were utilized for severity rating in tea leaf blight(TLB), while faster R-CNN was used to identify TLB leaves [12]. A Spatial attention and channel module integrated with the ResNet, to identify and categorize plant disease severity using an attention network model (SEV-Net) [13].

A novel lightweight CNN architecture, designated L-CSMS, was employed for plant disease severity detection on the PlantVillage dataset [14]. In another work, a two-stage model named DUNet was used on cucumber leaves on complicated backdrops by fusing DeepLabV3+ and U-Net [15, 16]. Fenu et al. [17] used five pretrained CNN architectures, and a self-collected field dataset called DiaMOS Plant was utilized. When the model was used to automatically and accurately extract the distinguishing characteristics of sick leaves, it showed resilience. Yin et al. [18] created DISE-Net and used a dataset of 1268 maize tiny leaf spots. While Liu et al. [19] used DeeplabV3+, PSPNet, and UNet models, and used a 5382 apple leaf images dataset, the purpose of both studies was to classify the disease severity. Using RGB image data from leaves affected by five plant diseases, used by Alves et al. [20] employed boosted regression tree models. A unique technique that reduces computing costs while optimizing the value of current data for better plant disease diagnosis accuracy. The program localizes leaf areas, performs color modification, and fits curves using RANSAC using the PlantVillage dataset [21]. A HEFL [22] and Fed-FSL [23] frameworks were used to detect tomato leaf diseases and cauliflower leaf diseases, respectively.

Russell et al. [24] used learnable filters and customized filters (Law's Mask) in a parallel multiscale stream to identify plant leaf diseases, while the MULTINET framework combines Multi-agent DRL with EfficientNet for 3D plant leaf disease detection and severity assessment. The method comprises AWW filter image pre-treatment, EMMARO data augmentation, BDM 2D to 3D transformation, EnDeep dual segmentation, and EBBI illness classification [25]. Lv et al. [26] used spectral bands (SBs), spectral indices (SIs), wavelet features (WFs), and first-order spectral derivatives (FSDs) to monitor Southern corn leaf blight (SCLB) severity classification. An automated methodology, employing image processing and deep learning convolutional neural networks (CNNs), was used for tomato plant diseases [27]. Yao et al. [28] used an integrated framework to segment target leaves

and disease spots in cucumber images for severity grading. It uses a U-Net to segment lesions with a Dice coefficient and TRNet to segment leaves.

Pal et al. [29] outline the use of the Agriculture Detection (AgriDet) framework, which uses Kohonen-based deep learning networks with a traditional Inception-Visual Geometry Group Network (INC-VGGN). A hybrid technique that uses CNN and bidirectional long-short-term memory (BiLSTM) was presented and used with unmanned aerial system (UAS) photos. [30]. An ensemble stacked deep learning model was presented by Gautam et al. [31] as a method for identifying and categorizing illnesses of mango leaves. To diagnose tomato leaf diseases, a Multi-Level Feature Fusion (MX-MLF2) model based on Modified-Xception was presented. To improve prediction accuracy, multi-level feature extraction and fusion were used [32]. A combination of a transfer learning model along with classical deep learning was used to classify the rice leaf diseases [33]. Capsule Network (CapsNet) for Vitis vinifera (grape) leaves leaf disease classification was developed [34].

DMCNN, a deep learning base approach, was presented to categorize tomato leaf diseases into many classes [35]. Using Optimized K-Means Clustering (OKMC) with SD-GHHO, the model extracts pictures from standard and real-time datasets, preprocesses them for improvement, and divides disease spots. CNN, VGG16, and ResNet are used to get the features, which are then concatenated using SD-GHHO [36]. A deep learning method named DFN-PSAN was used to diagnose plant diseases in the real environment. To improve accuracy, the model combines pyramidal squeezed attention (PSA) [37]. Existing methods are divided into four categories by the study: multi (model, label, output, and task). Three datasets are used to evaluate the proposed Generalized Stacking Multi-output CNN (GSMo-CNN) model: Plant Village, Plant Leaves, and PlantDoc [38].

Zhao et al. [39] introduced PDSNets, lightweight segmentation models for field-based plant disease severity assessment, achieving high accuracy and significantly improved prediction speed compared to LinkNet. It provides publicly available datasets for soybean, wheat, and apple disease data, enabling efficient and automated disease severity quantification. In another study, a device that enables real-time, pre-symptomatic disease detection in tomato plants, facilitating timely agricultural interventions, was introduced [40]. A ResNet50-based solution for mango fruit diseases, achieving high accuracy and deploying it in a mobile app for early detection in Sahelian regions [41]. A similar deep learning-based approach was used for rice disease severity detection [42]. SugarViT, a Vision Transformer model, for automated large-scale plant trait annotation and disease severity scoring in sugar beet, was used. The framework is designed to be generic, applicable to diverse image-based tasks, and demonstrates improved performance compared to state-of-the-art methods [43].

Deep document clustering and bibliometric analysis were used to explore the convergence of commercial strategies, sustainability, and AI-driven technologies in global food systems [44]. The analysis reveals that the integration of deep learning and optimization is essential for addressing macroeconomic demand and supply chain challenges, providing a strategic roadmap for interdisciplinary innovation in a sustainable global business environment. A lightweight YOLOv11 model for peanut leaf spot detection [45], YOLOv11-AIU [46], to grade tomato early blight severity. These studies used enhanced multi-scale feature extraction methodologies to reduce the computational complexity. In another study, YOLOv11-CARAFE-SE [47] with OpenCV-based HSV segmentation for area-based severity grading and YOLOv11-seg-DEDB [48], an optimized model for aloe anthracnose detection, proved 28% lighter than any other model, which was a significant achievement. Various other studies were used to enhance the feature extraction process, like the RSL Linked-TransNet model, which utilizes hierarchical feature extraction and transformer-based context modeling to perform multiclass segmentation of co-occurring citrus diseases [49], and a deep saliency map-based CNN framework that utilizes EfficientNet-B7 and Hybrid Harris Hawks Optimization to detect and assess the severity of seven maize leaf diseases [50]. Also, a crop-conditional semantic segmentation architecture that integrates contextual metadata and semi-supervised pseudo-labeling to improve disease detection across diverse field conditions [51]. Table 1 summarizes the comparative analysis of the most prominent literature work studies during this research.

The extensive review of related work reveals a clear trend towards the application of deep learning methodologies, predominantly various CNN architectures and increasingly Vision Transformers, for plant disease detection and severity assessment. Researchers are consistently enhancing model performance through architectural modifications. Despite advancements, current research still faces shortcomings. Many models struggle with generalizability to diverse environmental conditions and real-world complexities beyond controlled datasets. There's a persistent need for robust methods that can accurately assess multi-disease severity on a single leaf, especially when diseases overlap. Furthermore, improving computational efficiency for deployment on edge devices remains a critical area for further optimization.

Table 1: Comparative Analysis of the most prominent literature work

Source	Techniques	Strength	Weakness
Gunder, Maurice 2025 [43]	SugarViT	Efficient and accurate	Computational complexity
Alfred, Reuben 2025 [42]	DL Based	Generalizability and real-world applicability	Lacking nuanced insights
Faye, Demba 2025 [41]	ResNet50-based	High accuracy	Limitation to one crop

Zhao, Chenyi 2025 [39]	PDSNets	High accuracy	Lack of environmental condition
Yao, Jianping 2024 [36]	GSMo-CNN	Effective in handling multiple prediction tasks, achieves state-of-the-art performance.	May require more computational resources compared to simpler models.
Dai, Guowei 2024 [35]	DFN-PSAN	High accuracy, parameter efficiency, and improved transparency.	Complexity may increase computational cost.
Elfatimi, Elhoucine 2024 [33]	DMCNN	High accuracy, effective multi-scale feature learning.	Potentially high computational cost.
Abdalla, Alwaseela 2024 [52]	CNN-BiLSTM with domain adaptation	High classification accuracy, effective use of spatiotemporal data	Potential computational complexity
Pal, Arunangshu 2023 [53]	AgriDet framework	Improved accuracy, effective segmentation, and reduced overfitting	Potentially high computational cost
Yang, Bin 2023 [54]	TSTC network	High accuracy, effective multitask classification, and improved feature extraction	Complexity in implementation, potential overfitting due to multitask learning
Yao, Hui 2023 [26]	TRNet and U-Net	High segmentation accuracy, effective severity grading	Moderate Dice coefficient for lesion segmentation
Chelloug 2023 [55]	MULTINET Framework	High accuracy and precision	Complexity in implementation and computationally intensive

3. Proposed Methodology

Our proposed methodology combines the latest deep learning model with innovative approaches in response to the needs in multi-disease severity estimation on a single leaf in Pome crops. The architecture of the proposed methodology is shown in Figure 1, which presents a high-level flowchart of the proposed framework, summarizing the complete pipeline from raw image input to multi-disease severity output. It consists of three parts: the pre-processing, segmentation, feature extraction, and classification.

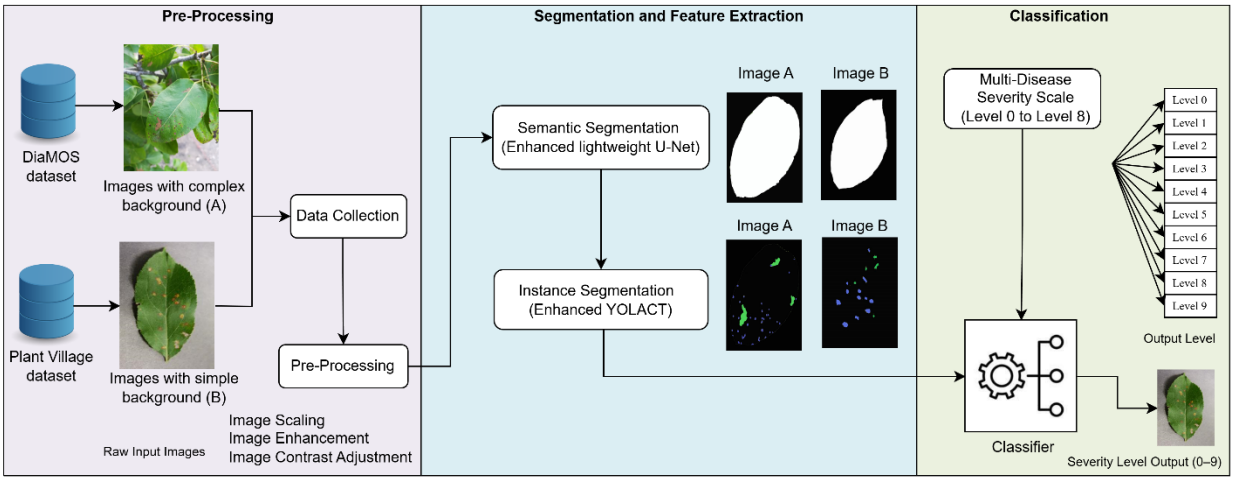

Figure 1. High-level flowchart of the proposed multi-disease severity estimation framework showing preprocessing, segmentation, disease detection, and final severity prediction.

3.1. Architecture Design Rationale and End-to-End Formalization

The selection of Lite-U-Net and Lite-YOLACT is motivated by complementary task specialization and computational efficiency considerations. Lite-U-Net provides encoder–decoder structured spatial reconstruction, which is analytically suitable for pixel-level lesion boundary recovery under irregular disease shapes. In contrast, Lite-YOLACT employs prototype mask generation with coefficient prediction, which enables instance-aware mask composition with reduced per-instance computation. From a computational perspective, encoder–decoder segmentation networks scale approximately with O(HW·C) feature operations per stage, whereas prototype-based instance segmentation reduces per-object mask generation to linear coefficient combination, lowering inference overhead compared to per-instance mask heads. The combined pipeline therefore achieves boundary precision and instance discrimination with lower parameter and FLOP growth than monolithic multi-task networks. This division of responsibilities yields a favorable accuracy–complexity tradeoff aligned with lightweight deployment objectives.

Let $I \in R^{H\times W\times 3}$ denote an input RGB leaf image. The proposed framework follows a three-stage end-to-end pipeline comprising segmentation, feature extraction, and classification. In the first stage, a segmentation network S(·) is applied to isolate disease-affected regions as shown in equation (1):

$$M = S(I) \tag{1}$$

Where $M \in \{0,1\}^{H\times W}$ represents the predicted disease mask.

In the second stage, the segmented image $I_8 = I \odot M$ is passed to a feature extraction module F(·), which learns discriminative representations as shown in equation (2):

$$f = F(I_8) \quad (2)$$

where $f \in R^d$ denotes the extracted feature vector.

Finally, the classification module C(·) maps the feature representation to a multi-disease severity level as shown in equation (3):

$$\hat{y} \in C(f) \quad (3)$$

where $\hat{y} \in \{0, 1, ..., 9\}$ corresponds to the defined multi-disease severity scale.

All modules are optimized sequentially and integrated to form a unified end-to-end severity estimation framework. Table 2 provides a layer-wise summary of the proposed framework, explicitly listing input and output dimensions for preprocessing, semantic segmentation, instance segmentation, and classification stages.

Table 2. Layer-Wise Architecture Details of the Proposed Framework

(a) Preprocessing			
Stage	**Operation**	**Input Size**	**Output Size**
Scaling	Resize	Raw Image	256 × 256 × 3
Enhancement	Contrast Adjustment	256 × 256 × 3	256 × 256 × 3
Labeling	Mask / BBox Annotation	256 × 256 × 3	Ground Truth Masks
(b) Semantic Segmentation – Lightweight U-Net (MobileNetV2 Backbone)			
Input	RGB Image	256 × 256 × 3	256 × 256 × 3
Encoder Block 1	MobileNetV2 (block_1)	256 × 256 × 3	128 × 128 × 32
Encoder Block 2	MobileNetV2 (block_3)	128 × 128 × 32	64 × 64 × 64
Encoder Block 3	MobileNetV2 (block_6)	64 × 64 × 64	32 × 32 × 96
Encoder Block 4	MobileNetV2 (block_13)	32 × 32 × 96	16 × 16 × 160
Bottleneck	MobileNetV2 (block_16)	16 × 16 × 160	8 × 8 × 320
Decoder	Upsample + DW Conv	—	256 × 256 × 64
Output Layer	1×1 Conv + Sigmoid	256 × 256 × 64	256 × 256 × 1
(c) Instance Segmentation – Lite-YOLACT			
Backbone	MobileNetV2 + FPN	256 × 256 × 3	Multi-scale Features
Protonet	Prototype Mask Generation	FPN Features	k × H × W
Prediction Head	BBox + Coefficients	FPN Features	N × (Boxes + Classes + Coeffs)
Mask Assembly	Linear Combination	k Masks + Coeffs	Instance Masks
Output	Sigmoid Activation	—	Binary Instance Masks
(d) Classification Head			
Feature Vector	Flattened Segmentation Features	d	d
FC Layer 1	Dense + ReLU	d	512
Dropout	p = 0.5	512	512
FC Layer 2	Dense + Softmax	512	10 (Severity Levels)

3.2. Preprocessing

The proposed methodology begins with a comprehensive preprocessing phase to ensure consistent and optimized input for model training. All raw

images were uniformly resized to a standard resolution of 256 × 256 × 3 to maintain dimensional consistency required by deep learning architectures. Pixel normalization was subsequently applied to standardize input values, stabilizing and accelerating the training process. Contrast enhancement techniques were also employed to improve lesion visibility and accentuate subtle structural variations within leaf regions. To reduce background complexity, preliminary background removal techniques such as thresholding and edge detection were applied to generate initial leaf masks. These steps allow the model to focus on biologically relevant regions during segmentation and classification. The dataset was annotated using the open-source tool [56] to generate precise bounding boxes and pixel-wise segmentation masks. Annotation was performed by three certified plant pathology experts with extensive experience in pome crop disease assessment. Severity levels (0–8) were assigned based on quantitative pixel-level lesion area estimation, lesion density distribution, and established agricultural pathology guidelines. In cases of disagreement, a majority voting scheme was applied to minimize subjectivity. Inter-annotator reliability was measured using Cohen's Kappa coefficient ($\kappa = 0.91$), indicating strong agreement. Severity thresholds and level boundaries were further reviewed using domain guidelines and validated through inter-class separability analysis during dataset preparation.

To improve generalization, extensive data augmentation was applied during training, including rotation, flipping, scaling, and translation. These transformations enable the model to learn variations in leaf orientation, appearance, and environmental conditions. The dataset exhibits class imbalance across disease severity levels, with mild and moderate cases being more prevalent. To mitigate learning bias, class-aware augmentation was applied to underrepresented classes to increase effective sample diversity. Additionally, class weights were incorporated into the loss function to penalize misclassification of minority classes more heavily. These strategies collectively promote balanced learning and improve robustness across all disease categories and severity levels.

3.3. Segmentation using Lightweight U-Net

To enhance the performance and efficiency of the traditional U-Net [57] architecture for semantic segmentation, we propose a lightweight modification that significantly reduces computational complexity while keeping up the highest accuracy. The improved design replaces standard convolutional layers with a depth-wise and a pointwise convolution. This change drastically reduces the number of parameters, thus floating-point operations (FLOPs). We have also replaced transposed convolutions in the decoder with bilinear up-sampling followed by convolution, further accelerating inference without compromising detail recovery. The encoder-decoder structure remains intact, preserving skip connections for effective multi-scale feature fusion. Optional integration of pretrained lightweight

backbones such as MobileNetV2 can further enhance the feature extraction capability while keeping the model compact. These modifications collectively made our Lite-U-Net architecture acceptable to be implemented in real-time environments such as mobile and other devices, without sacrificing segmentation accuracy. Details of the improved Lightweight U-Net with MobileNetV2 Backbone are shown in Algorithm 1.

Semantic segmentation using U-Net is employed to ensure that only the leaf is selected, and everything else in the background is removed. U-Net is a convolutional neural network designed for biomedical image segmentation, characterized by its U-shaped architecture. First, the image is passed through the U-Net, where the encoder path captures the context by down-sampling the image and the decoder path reconstructs the segmentation mask by up-sampling it. Fine-grained details are preserved because of skip connections between layers in the encoder and decoder. The output is a binary mask where pixels belonging to the leaf are labeled as 1 and the background as 0. This mask is applied to the original image to obtain a final segmented image with only the leaf remaining. Figure 2 shows the layers block diagram of the Lightweight U-Net with MobileNetV2 Backbone.

Algorithm 1: Lightweight U-Net with MobileNetV2 Backbone
Input: RGB image of size 256×256×3
Output: Segmentation mask
Encoder (Feature Extractor):
 a. Use Pretrained MobileNetV2 model as encoder
 b. Extract multi-resolution feature maps from selected layers (e.g., block_1, block_3, block_6, block_13, block_16)
 c. Downsampling by strides in MobileNetV2
Bottleneck:
 a. Use the last layer of MobileNetV2 as bottleneck
Decoder (Lightweight Expanding Path):
 a. Apply UpConvolution (Upsample + Conv)
 b. Concatenate with corresponding encoder feature map
 c. Apply Depthwise Separable Convolution + ReLU
 d. Repeat for all decoder levels
Final Layer:
 a. Apply 1×1 Conv to reduce to number of classes
 b. Apply Softmax activation

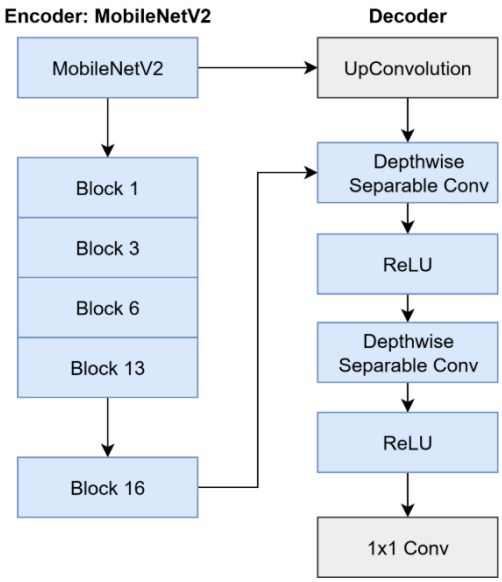

Figure 2. Architecture of the lightweight U-Net model with a MobileNetV2 backbone used for leaf segmentation.

Once the Lite-U-Net model has been trained, post-processing steps are applied that refine the segmentation results. During this step, a binary mask is created by applying a threshold that distinguishes the leaf from the background. To further refine the mask, morphological dilation and erosion are applied, which remove noise to ensure a smooth and accurate representation of the leaf. At certain places where leaf edges were not cleared, we have employed a contour detection technique. For evaluation, quantitative metrics, like IoU, Dice coefficient, and pixel accuracy, are used. Additionally, a visual inspection was also performed to ensure the model's effectiveness. This comprehensive evaluation process ensures that the semantic segmentation is accurate and meets the desired criteria for further analysis.

3.4. Segmentation with YOLACT

After the semantic segmentation, we have employed YOLACT [58] for robust instance segmentation. YOLACT is a cutting-edge, real-time instance segmentation model that distinguishes itself by decoupling the complex task into two parallel, more efficient branches: the generation of a set of class-agnostic prototypical masks and the prediction of instance-specific mask coefficients. Its architecture typically leverages a powerful backbone MobileNetV2 with FPN to extract rich features, which then feed into these two branches. The final instance masks are then efficiently produced by linearly combining the predicted coefficients with the global prototypes. This innovative design grants YOLACT significant advantages, including real-time processing speeds and superior computational efficiency, all while maintaining competitive accuracy, making it ideal for our subsequent, fine-grained segmentation requirements. Figure 3 shows the block diagram of our enhanced Lite-YOLACT model.

For a given input image *I*, YOLACT works in the following way. Feature Extraction using a backbone network and FPN extracts multi-scale features, represented as *F*. The Protonet generates a set of k class-agnostic prototype masks, denoted as $P=\{p_1,p_2,\ldots,p_k\}$, where each p_i is an H×W mask. The Prediction Head, based on the extracted features *F*, predicts *N* detections (bounding boxes and class scores) and, for each detection *j*, a vector of *k* mask coefficients $Cj=\{cj1,cj2,\ldots,cjk\}$. For each detected instance *j*, its final instance mask M_j is computed by a linear combination of the prototype masks P and its corresponding coefficients C_j, followed by a sigmoid activation σ as shown in equation 4.

$$M_j = \sigma\left(\sum_{i=1}^{k} C_{ji} * p_i\right) \tag{4}$$

Where:

- M_j is the binary mask for the j-th detected instance.
- c_{ji} is the i-th mask coefficient for the j-th instance.
- p_i is the i-th prototype mask.
- σ is the sigmoid activation function, which squashes the values to a range between 0 and 1, allowing for thresholding to create a binary mask.

After training our Lite-YOLACT, post-processing steps are applied, which refine the segmentation results. During this step, a binary mask is created by applying a threshold that distinguishes the leaf from the background. In the case of overlapping regions, Non-Maximum Suppression (NMS) is utilized to get the most accurate bounding boxes and masks, ensuring that the multiple detections are handled effectively. To further refine the mask, morphological dilation and erosion are applied, which remove noise to ensure a smooth and accurate representation of the leaf. For evaluation, quantitative metrics, like Average Precision (AP), mean Intersection over Union (mIoU), and mask quality, are used. Additionally, a visual inspection was also performed to ensure the model's effectiveness. This comprehensive evaluation process ensures that the semantic segmentation is accurate and meets the desired criteria for further analysis. ROC-AUC values were computed using macro-averaged multi-class ROC curves and averaged over 20 independent runs to ensure robustness.

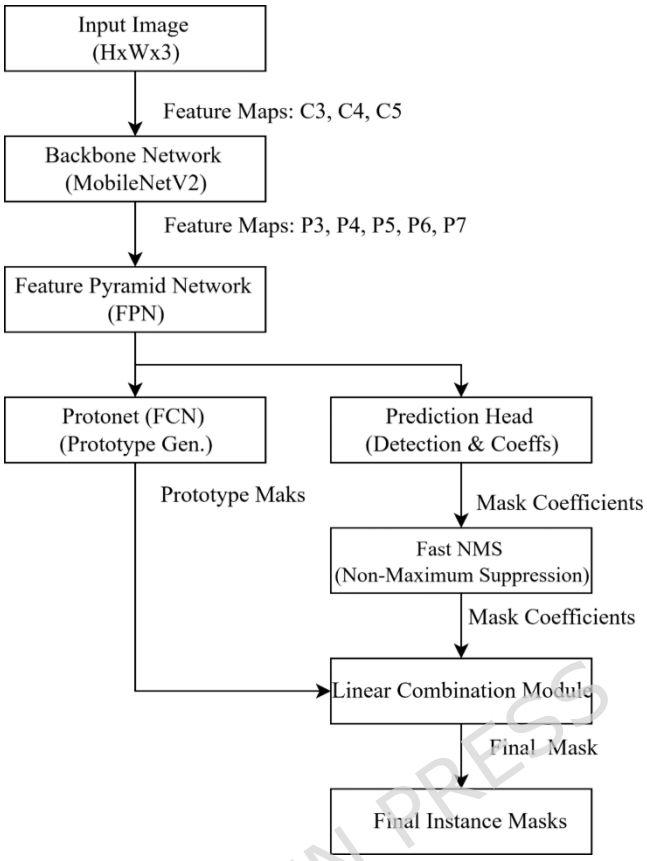

Figure 3. Architecture of the lightweight YOLACT model used for disease detection and instance segmentation.

3.4.1. Loss Function:

For optimizing the YOLACT model, we employ a comprehensive multi-task loss function of instance segmentation. Mask Loss (Lmask) [59] is the core component for pixel-level instance segmentation. Binary Cross-Entropy (BCE) is used for the linearly combined prototype masks. For each detected instance, the BCE loss is calculated pixel-wise between the predicted instance mask (after sigmoid activation) and its corresponding ground truth binary mask. This loss ensures that the model accurately predicts the presence or absence of an object's pixels within its segmented region, as shown in equation 5.

$$L_{total} = \beta L_{mask} \quad (5)$$

In our experiments, the weighting parameter β was set to **0.6**, giving slightly higher emphasis to mask prediction compared to bounding box regression after preliminary tuning. The value was selected through grid search over {0.3, 0.5, 0.6, 0.7, 1.0} on the validation set. Performance variation remained within ±0.8% across this range, indicating stable behavior, while β = 0.6 provided the best validation F1 score. The coefficient β is a hyperparameter that controls the relative weighting of the

mask loss components, allowing us to balance their contributions to the overall training objective. This multi-task loss function enables YOLACT to learn all necessary aspects of instance segmentation effectively and simultaneously.

3.5. Formal Definition of Multi-Disease Severity Scale

A multi-disease severity level scale is proposed to assess the two diseases on a single plant leaf. As per the proposed scale, severity levels are 0-8, which means 0 is the lowest when both diseases are < 10%. If both diseases are less than 50%, then it is level 1, and if both diseases are less than 100%, it is the end stage of the disease. These are the cases when the difference between disease pixels of both diseases is < 10%. Similarly, for Level 3, Disease 2 should be less than 10%, and Disease 1 should be less than 25% but greater than 10%, which will make Disease 1 at the Mid stage, and Disease 2 at the initial stage. In level 4, Disease 2 should be less than 10%, and Disease 1 should be less than 50% but greater than 10%. Disease 1 is at the Final stage, disease 2 is at the initial stage. In level 5, Disease 1 should be less than 10%, and Disease 2 should be less than 25% but greater than 10%. Disease 2 is at mid-stage, and disease 1 is at the initial stage. In level 6, Disease 1 should be less than 10%, and Disease 2 should be less than 50% but greater than 25%. Disease 2 is at the Final stage, disease 1 is at the initial stage. In level 7, Disease 2 should be less than 50%, and Disease 1 should be less than 100% but greater than 50%. Disease 1 is at the Final stage, disease 2 is at the Mid stage. In level 8, Disease 1 should be less than 50%, and Disease 2 should be less than 100% but greater than 50%. Disease 2 is at the Final stage, and Disease 1 is at the Mid stage. Details of the Multi-Disease severity level scale are given below in Table 3.

Table 3: Multi-Disease Severity Scale based on normalized lesion-area ratios and agronomic severity grading thresholds

Severity Levels	Multi-Disease Scale	Scale Description
Level 0	Disease 1, Disease 2 < 10%	Both Diseases are at initial stage
Level 1	10% < (Disease 1, Disease 2) < 25%	Both Diseases are at Mid stage
Level 2	25% < (Disease 1, Disease 2) < 50%	Both Diseases are at End stage
Level 3	50% < (Disease 1, Disease 2) < 100%	Both Diseases are at End stage
Level 4	Disease 1 < 10% and 25% > Disease 2 > 10% (internal difference > 10%)	Disease 2 is at Mid stage, Disease 1 at Initial stage
Level 5	Disease 1 < 10% and 50% > Disease 2 > 25% (internal difference > 10%)	Disease 12 is at Sever stage, Disease 1 at Initial stage
Level 6	Disease 1 < 10% and 100% > Disease 2 > 50% (internal difference > 10%)	Disease 2 is at End stage, Disease 1 at Initial stage

Level 7	10% < Disease 1 < 25% and 50% > Disease 2 > 25% (internal difference > 10%)	Disease 2 is at Sever stage, Disease 1 at Mid stage
Level 8	10% < Disease 2 < 25% and 100% > Disease 1 > 50% (internal difference > 10%)	Disease 2 is at End stage, Disease 1 at Mid stage
Level 9	25% < Disease 1 < 50% and 100% > Disease 2 > 50% (internal difference > 10%)	Disease 2 is at End stage, Disease 1 at Sever stage

To ensure reproducibility and reduce ambiguity, the proposed multi-disease severity scale is formally defined using normalized lesion-area ratios derived from segmentation masks. Let A_L denote the total leaf area and A_{d_i} denote the pixel area corresponding to disease *i*. The disease severity ratio is defined in equation (6):

$$S_{d_i} = \frac{A_{d_i}}{A_L} \tag{6}$$

Severity levels are assigned by discretizing S_{d_i} into bounded intervals. The selected thresholds (<10%, 10–25%, 25–50%, >50%) follow commonly used agronomic visual grading ranges reported in plant pathology severity assessment practices. These intervals provide a balance between pathological interpretability and statistical separability across severity classes. For multi-disease cases, the final severity level is determined using joint threshold rules and inter-disease dominance constraints, enabling consistent labeling when multiple diseases coexist on the same leaf.

3.6. Classification

Following the precise semantic and instance segmentation achieved by our Lightweight U-Net and YOLACT models, the extracted and refined visual information is then channeled into a Fully Connected (Dense) Neural Network for robust classification. This classifier is strategically positioned to leverage the rich, high-level features derived from the segmented leaf and its individual disease instances. By flattening these powerful feature representations into a vector, the dense network can learn intricate, non-linear relationships crucial for accurate categorization. This architecture allows for flexible integration of various derived metrics, such as disease percentages or spatial distributions, enabling the model to precisely classify specific disease types, assess their severity levels, or even identify complex multi-disease states, thereby completing the analytical pipeline with a highly discriminative decision-making component. Figure 4 shows the proposed classification hierarchical approach.

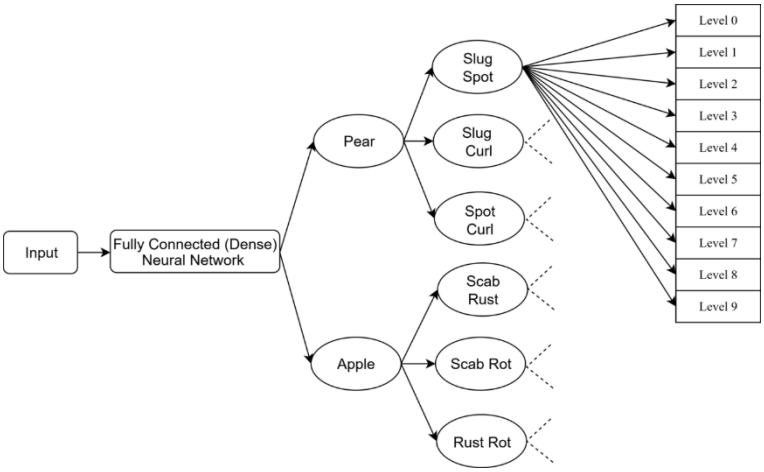

Figure 4: Proposed hierarchical classification approach for multi-disease severity estimation across different disease categories and severity levels

4. Experiment Results

This section explains the dataset, preprocessing details, and the evaluation metrics employed in this research.

4.1. Dataset

The datasets employed in this research, including the DiaMOS [60] and PlantVillage [61] datasets, are publicly accessible. The PlantVillage dataset consists of annotated images of healthy and diseased crop leaves, with labels provided for specific disease types and severity levels. Similarly, the DiaMOS dataset comprises high-resolution images of Pome leaves with multi-disease annotations. Figure 5 shows a sample of images from our datasets.

Figure 5: Sample pear leaf images from the DiaMOS dataset and apple leaf images from the PlantVillage dataset used in this study.

This study investigates the potential of plant disease severity in the diagnosis of Pome fruit plants, Apple, and Pear only. Pear images were taken from the DiaMOS dataset. This dataset contains 4 classes of diseases, like curl, slug, and spot, and each class has five different stages (Level-0 to

Level-4). Here, Level-0 belongs to low severity, and Level-4 belongs to the severe category. The PlantVillage dataset was used for apple leaf images. Which contains four disease classes: healthy, rust, scab, and black rot and each class was divided into 5 severity stages (Level-0 to Level-4). Multiple diseases on a single leaf were found and divided into 6 categories: Slug Spot, Slug Curl, Spot Curl for Pear dataset, and Scab Rust, Scab Rot, and Rust Rot for Apple leaves. Although each disease category comprised approximately 400 images, details of the dataset are given below in Table 4.

Table 4: Data Augmentation

Description	Images	Images After Augmentation
Slug Spot - Pear	400	1600
Slug Curl - Pear	350	1400
Spot Curl - Pear	380	1520
Scab Rust - Apple	400	1600
Scab Rot - Apple	350	1400
Rust Rot - Apple	360	1440

Table 5 summarizes the dataset composition and experimental split used in this study. The dataset consists of 8960 images covering three diseases with three severity levels each. The data were divided into training (70%), validation (15%), and testing (15%) subsets at the class level to avoid data leakage and ensure fair performance evaluation.

Table 5. Dataset Statistics and Experimental Split

Disease Class	**Severity Level**	**Total Images**	**Training (70%)**	**Validation (15%)**	**Testing (15%)**
Slug Spot – Pear	Initial	400	280	60	60
	Mild	400	280	60	60
	Moderate	400	280	60	60
	Severe	400	280	60	60
Slug Curl – Pear	Initial	350	245	53	53
	Mild	350	245	52	52
	Moderate	350	245	52	52
	Severe	350	245	52	52
Spot Curl – Pear	Initial	380	266	57	57
	Mild	380	266	57	57
	Moderate	380	266	57	57
	Severe	380	266	57	57
Scab Rust – Apple	Initial	400	280	60	60
	Mild	400	280	60	60
	Moderate	400	280	60	60
	Severe	400	280	60	60
Scab Rot – Apple	Initial	350	245	52	52
	Mild	350	245	52	52
	Moderate	350	245	52	52

	Severe	350	245	53	53
Rust Rot – Apple	Initial	360	252	54	54
	Mild	360	252	54	54
	Moderate	360	252	54	54
	Severe	360	252	54	54
Total	**2,240**	**8,960**	**6,272**	**1,344**	**1,344**

4.2. Experimental settings and metric used:

The proposed system was implemented in Python, leveraging GPU acceleration for deep learning operations. Experiments were conducted on a system equipped with an Intel Quad-Core i7 2820QM CPU (2.30 GHz), an NVIDIA Quadro GPU, and a Samsung 850 Pro 256 GB SSD for both training and testing. This study evaluated the proposed model using standard metrics: accuracy, precision, recall, and F1-score, and Area Under the Receiver Operating Characteristics (AUC-ROC). These metrics are calculated based on True Positives (TP, correctly classified diseased images), True Negatives (TN, correctly classified healthy images), False Positives (FP, healthy images classified as diseased), and False Negatives (FN, diseased images classified as healthy). To ensure model generalization and statistical robustness, a 5-fold cross-validation strategy was employed, where the dataset was partitioned into five equal subsets with a 4:1 training-to-testing ratio per fold. This process was repeated across 20 independent runs using distinct random initialization seeds and shuffled dataset splits. Final performance metrics are reported as the Mean ± Standard Deviation, supplemented by 95% confidence intervals to quantify stability across all folds and runs. This rigorous evaluation framework ensures that the reported results reflect consistent performance rather than artifacts of specific data partitioning.

4.3. Semantic Segmentation Results

Image semantic segmentation results of the proposed model on the DiaMOS and PlantVillage datasets are discussed in this section. For this purpose, multi-diseased Pome leaves were used, specifically focusing on apple and pear leaves. The proposed approach effectively addresses these challenges, accurately detecting diseases and their percentage even in images with complex backgrounds and different positional variations. Figure 6 shows the images with the region of interest selected and the segmented output using Lite-UNet from the sample images.

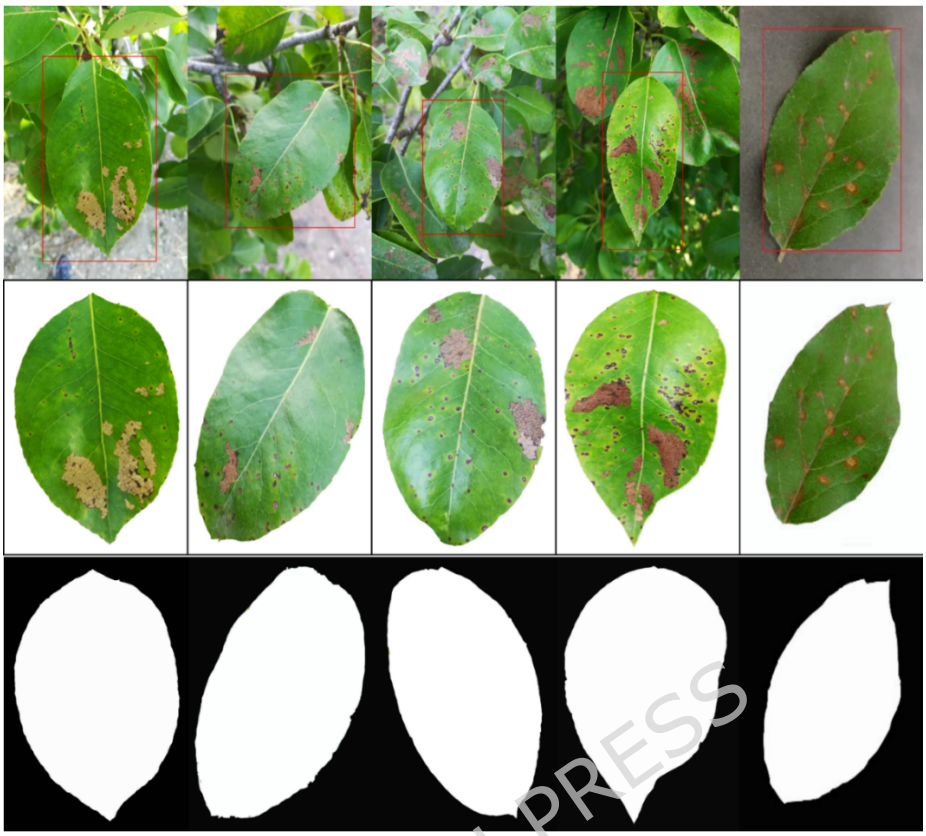

Figure 6: Example leaf segmentation results generated by the Lite-U-Net model.

To unequivocally underscore the enhanced efficacy of our proposed Lite-UNet, a meticulous comparative analysis was undertaken against its conventional U-Net counterpart. Both architectures were subjected to identical training regimens and subsequently evaluated with rigorous impartiality on the same test dataset, leveraging the quantitative metrics. The precise configurations employed for training both models are systematically documented in Table 6. Beyond mere performance, a comprehensive assessment of computational complexity using the number of parameters, Floating-Point Operations (FLOPs), and inferential latency (ms) was performed. This comparative data, as presented in Table 7, robustly demonstrates the Lite-UNet's superior performance across all evaluated aspects, thereby affirming its inherent efficiency. Sensitivity analysis shows the reason of choosing loss weight value 0.6 as show in Table 8.

Table 6: Hyperparameter settings used for training all models in the proposed framework.

Hyperparameter	**Values**
Epochs	40
Learning rate	0.0001
Batch size	32

Learning Rate Scheduler	ReduceLROnPlateau
LR reduce factor	0.2
LR Patience	5 epochs
Optimizer	Adam
Loss function	Dice loss
β (loss weight)	0.6

Table 7: Comparison of Model Performance for classical U-Net with proposed Lite-UNet

Technique	Accuracy (%)	IoU	FLOPS	Dice	Parameters	Inference Time (ms) GPU
U-Net	90.36	76.0	1.09G	84.0	31M	98
Lite-UNet	95.12	85.0	0.041G	90.0	17.3M	57

Table 8: Sensitivity Analysis

β	**F1**
0.3	94.22
0.5	94.81
0.6	**95.12**
0.7	94.90
1.0	94.35

4.4. Instance Segmentation Results

This Section provides details on the training configurations and results derived from the application of the proposed Lite-YOLACT model for instance segmentation. The model's performance is rigorously assessed on both datasets and its corresponding instance mask annotations. Our proposed instance segmentation approach, leveraging Lite-YOLACT's real-time capabilities, is specifically engineered to operate efficiently on mobile and low-computing devices. This is achieved through its innovative architecture that decouples mask prediction, significantly reducing network latency and enabling seamless operation within real-time environments for fine-grained object delineation. Figure 7 shows the images with the region of interest selected and the segmented output from the sample images.

500 random samples from each dataset, DIAMOS and plant village datasets, when using one upsampling layer (2x upscaling) and when using two upsampling layers (4x upscaling). Because of the requirement of having a lightweight model, so that the application could be used in the filed real time environments, we also compared both models in terms of complexity (parameters and FLOPs) and latency (inference time) as shown in Table 9.

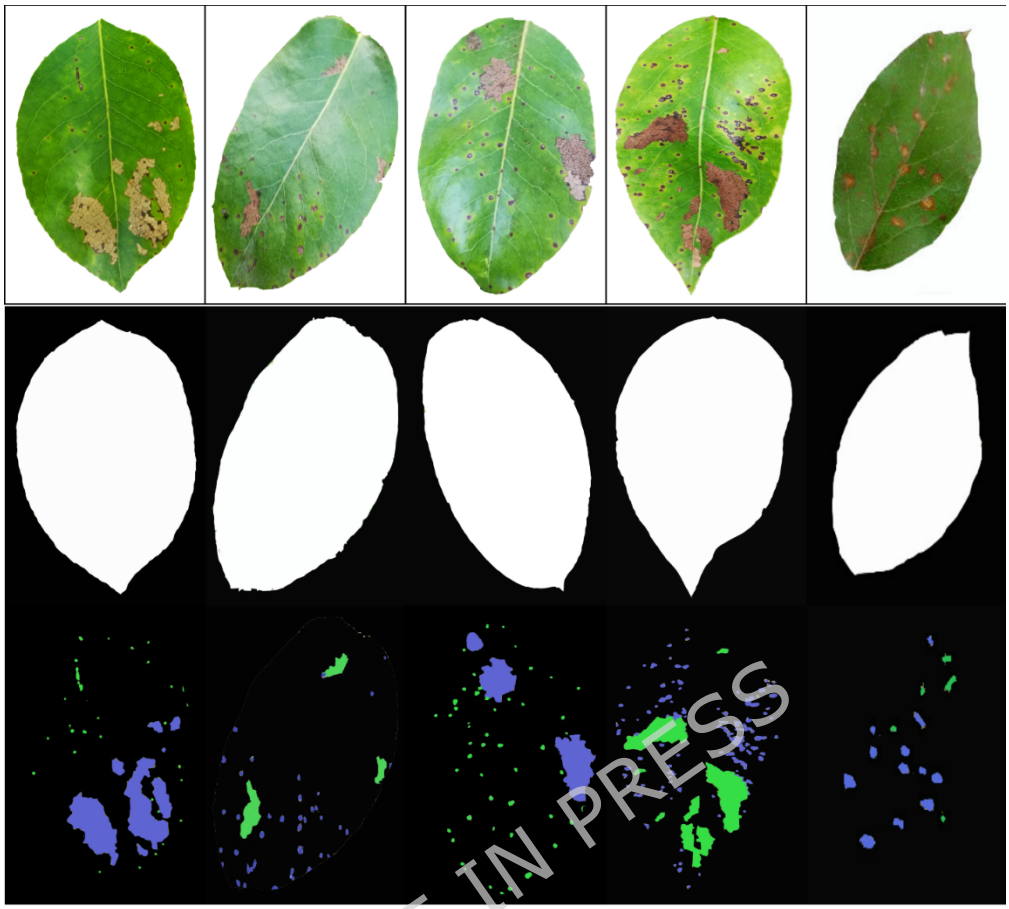

Figure 7: Disease detection and localization results produced by the Lite-YOLACT model

Table 9: Comparison of Model Performance for YOLACT and Lite-YOLACT

Upsampling	Method	Parameters	FLOPs	Average Inference Time (ms)
64p to 256p	YOLACT	44M	250	150
128p to 256p	Lite-YOLACT	15M	45	30

4.5. Effect of Loss Functions:

The Mask Loss (Lmask) component within YOLACT's multi-task loss function is pivotal for achieving precise pixel-level instance segmentation. Its primary effect is to guide the model in accurately classifying each pixel as either foreground or background, thereby directly influencing the refinement of object shapes and boundaries. By penalizing discrepancies between predicted and ground truth masks, Lmask ensures that the model learns to generate highly accurate contours, driving the optimization of both the class-agnostic prototype masks and the instance-specific coefficients. Ultimately, this loss is instrumental in enabling YOLACT to produce high-quality, distinct, and pixel-accurate segmentations for each

detected object, which is fundamental for downstream tasks requiring fine-grained spatial information.

4.6. Proposed Work Classification Results:

The proposed model underwent a training process on the selected dataset. The model demonstrated strong performance, with a training accuracy of 98% and a validation accuracy of 95%. This indicates the good generalization ability of the model. The training loss and validation loss both converged to the lowest value of 0.1, suggesting that the model demonstrated effective learning without exhibiting signs of overfitting. Figure 8 (a) and (b) show the training accuracy and validation accuracy, training loss, and validation loss over epochs of the proposed methodology.

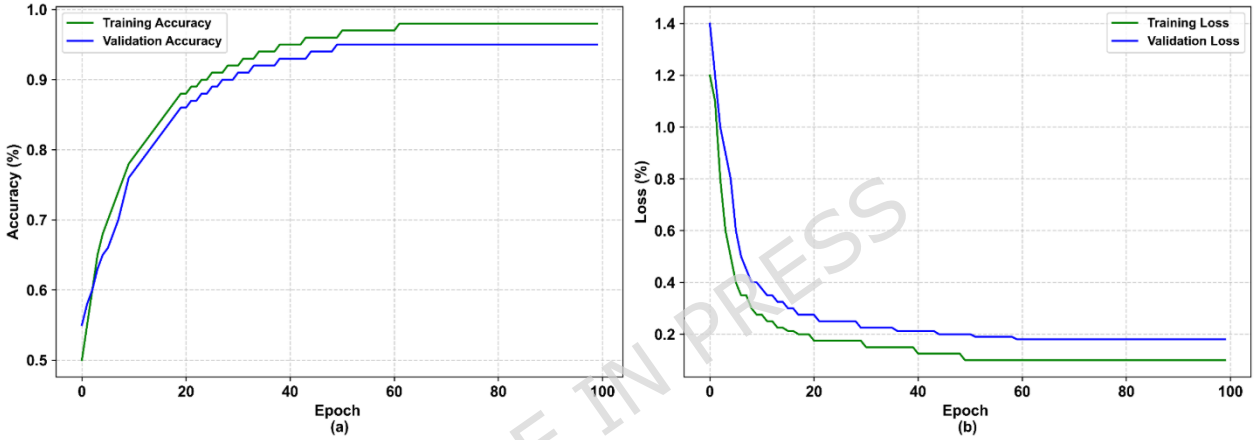

Figure 8: Training performance of the proposed model: (a) training vs validation accuracy and (b) training vs validation loss.

In our comprehensive evaluation, we have used Pome leaf images of multiple diseases on a single leaf, with diverse disease conditions and severity levels. Due to our fine-tuned MobileNetV2 backbone model for feature extraction, the proposed model was able to cross an accuracy of 95% in multiple disease symptoms. Other than accuracy, we have analyzed precision, recall, F1-score, and ROC-AUC metrics to evaluate our model, as shown in Table 10. Our proposed model exhibits high precision, which indicates a low rate of false positives and strong recall metrics, demonstrating its capability to identify disease across various severity levels. At the preprocessing stage, significant improvements have been seen in the results. At the start, the Lite-U-Net architecture effectively segmented the leaf images, separating diseased areas from the complex backgrounds.

Table 10. Per-Class Performance Analysis of the Proposed Model (20 Independent Runs)

Disease Class	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC	SD (ROC-AUC)
Slug Spot –	92.0	94.0	94.90	0.96	0.008

Pear					
Slug Curl – Pear	91.5	93.8	94.60	0.94	0.010
Spot Curl – Pear	92.8	94.2	95.00	0.95	0.009
Scab Rust – Apple	94.5	95.6	96.10	0.97	0.007
Scab Rot – Apple	91.0	93.5	94.50	0.93	0.012
Rust Rot – Apple	96.2	95.0	96.62	0.95	0.008
Average	**93.00**	**94.35**	**95.12**	**0.95**	**0.009**

The per-class analysis demonstrates that Scab Rust, Apple achieves the highest ROC-AUC (0.97) due to well-defined lesion boundaries and strong chromatic contrast. In contrast, Scab Rot, Apple exhibits relatively lower ROC-AUC (0.93) due to higher intra-class variability and feature overlap with adjacent severity levels. Results are averaged over 20 independent runs. The accurate segmentation allowed the model to focus on the ROI for feature extraction. Then the U-Net's encoder-decoder structure saved spatial information, necessary for accurate disease classification. The integration of Lite-YOLACT refined the segmentation process by predicting object masks, enhancing the model's ability to detect multiple disease instances on a single leaf and improving both recall and precision metrics. Lite-YOLACT's capability to generate detailed masks ensured that even subtle disease symptoms were correctly identified and classified.

Table 11 shows three examples of how the severity level is calculated using the multi-diseases severity scale. 6500 total leaf pixels out of which 2500 pixels are diseased, making 38%. Disease 1 pixels are 1700, which makes 26%, and 800 for disease 2, which makes 12%. The difference between the two diseases is 14%. As per the multi-disease scale, Level 7: Disease 1 is at the severe stage, and Disease 2 is at the Mild stage.

Table 11: Results calculation of the multi-disease leaf using the proposed scale

Total Leaf Pixels	Total Diseased Pixels	Disease 1 (D1) Pixels	Disease 2 (D2) Pixels	D1%	D2%	% Difference	D1, D2%	Results
6500	2500	1700	800	26	12	14	38	Level 7: Disease 1 is at Sever stage, Disease 2 at the Mild stage
6000	1000	700	300	11	5	4	16	Level 4: Disease 1 is at Mild stage, Disease 2 at Initial stage
7000	4000	3200	800	45	11	34	57	Level 7: Disease 1 is at Sever stage, Disease 2 at the Mild stage

4.7. Quantitative Validation of Grad-CAM Interpretability

The use of Grad-CAM further validates and visualizes the interpretability of our model's decisions. Produced results that provided insights into the model's decision-making process by producing heatmaps that visually indicate crucial spots on Pome leaf pictures that influence disease severity classification. For each disease spot, the severity is visually represented through color gradation: the center of the spot, indicating severe disease, is marked in red; the middle area, showing mild effects, is represented by orange; and the outer area, where the effects are minimal, is represented by yellow. Using Grad-CAM, we can quantify the severity by counting the number of pixels corresponding to each color (red, orange, and yellow) within a disease spot. This provides a detailed representation of the first disease. The same process can be applied to other disease spots, enabling a precise assessment of severity for each disease. Figure 9 shows a Heatmap of sample images using Grad-CAM.

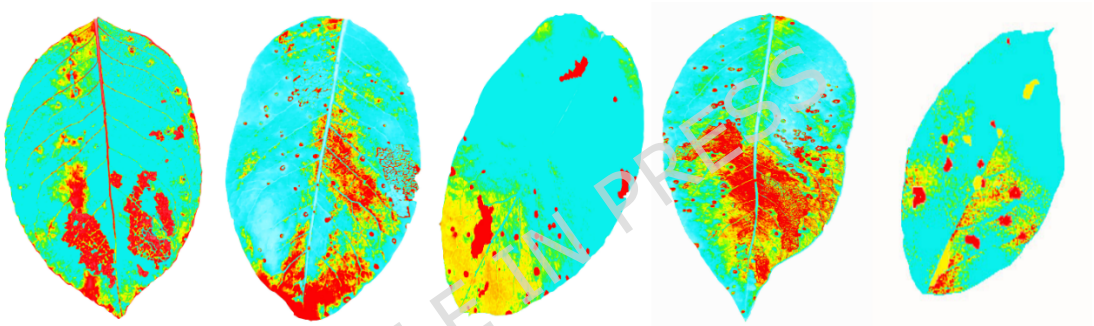

Figure 9: Grad-CAM heatmap visualization highlighting regions contributing to disease severity prediction

To quantitatively evaluate the alignment between Enhanced Grad-CAM heatmaps and ground truth segmentation masks, we computed spatial overlap metrics on a representative subset of 300 test images as shown in equation (7) and (8).

$$\text{IoU} = \frac{|G \cap H|}{|G \cup H|} \tag{7}$$

$$\text{Dice} = \frac{2|G \cap H|}{|G| + |H|} \tag{8}$$

Where, G = Ground truth mask and H = Thresholded Grad-CAM region.

The high Dice similarity (0.89) confirms that the network's attention mechanism accurately localizes diseased regions. Notably, lower IoU for Scab Rot corresponds to its irregular lesion morphology, consistent with the lower ROC-AUC observed earlier. This correlation further strengthens the interpretability-validity relationship as shown in Table 12.

Table 12: Quantitative Evaluation of Grad-CAM Alignment

Disease Class	IoU	Dice Score	Pointing Accuracy (%)
Slug Spot – Pear	0.83	0.88	91.5
Slug Curl – Pear	0.82	0.87	90.8
Spot Curl – Pear	0.85	0.89	92.1
Scab Rust – Apple	0.87	0.91	94.2
Scab Rot – Apple	0.80	0.86	89.4
Rust Rot – Apple	0.86	0.90	93.0
Average	**0.84**	**0.89**	**92.3**

4.8. Computational Efficiency Analysis

To evaluate the practical applicability of the proposed framework, we conducted a detailed computational efficiency analysis. Table 13 summarizes parameter count, FLOPs, memory consumption, training time, and inference latency on both GPU and CPU platforms. All models were trained and evaluated under identical conditions. The proposed method achieves a favorable balance between accuracy and computational cost, demonstrating suitability for real-time and edge-based multi-disease severity assessment. Compared with unified multi-task segmentation networks, the proposed dual lightweight architecture reduces prototype mask generation cost and encoder–decoder redundancy, resulting in lower FLOPs and faster inference.

Table 13. Computational Efficiency Analysis

Model	Parameters (M)	FLOPs (GFLOPs)	Memory Usage (MB)	Training Time (hrs/ epochs)	GPU Inference (ms/image)	CPU Inference (ms/image)
ResNet50	25.6	4.1	98.4	6.2 / 100	42.3	186.5
EfficientNet-B4	19.0	4.2	72.6	5.4 / 100	38.1	164.8
U-Net	31.0	5.6	118.2	7.1 / 100	45.7	201.3
YOLOv5	7.5	16.5	27.4	4.3 / 100	21.6	98.7
Proposed Method	**12.8**	**6.3**	**54.9**	**4.9 / 100**	**24.8**	**112.4**

4.9. Failure Case Analysis and ROC Curve Evaluation

While the proposed framework achieves strong overall performance, certain challenging scenarios were observed where performance degrades. One common failure case occurs when co-existing diseases exhibit high visual similarity, particularly at early or mid-severity stages, where texture and color variations are subtle. In such cases, instance segmentation may partially overlap disease regions, leading to inaccurate disease area estimation and severity assignment. The model is also sensitive to low-contrast images, uneven illumination, and shadows, which affect boundary precision during segmentation. Additionally, when infected regions occupy a very small portion of the leaf, the limited discriminative information can

result in an underestimation of disease severity. Complex leaf structures, including curling, occlusion, and overlapping leaves, further challenge instance separation, especially when one disease dominates the leaf surface.

To evaluate classification robustness beyond accuracy-based metrics, class-wise Receiver Operating Characteristic (ROC) curves were generated using a one-versus-rest approach as shown in Table 9. The ROC analysis demonstrates consistently high discriminative performance across most disease classes, with strong Area Under the Curve (AUC) values for diseases with distinct visual characteristics. Comparatively lower AUC values were observed for visually similar disease pairs and early-stage infections, which aligns with the identified failure scenarios. This analysis provides a clearer understanding of the model's limitations and highlights the need for improved robustness to illumination variability, fine-grained disease differentiation, and complex real-field conditions in future extensions. Figure 10 presents the class-wise ROC curves, showing consistently high AUC values for most disease categories, while comparatively lower separability is observed for visually similar disease pairs and early-stage infections.

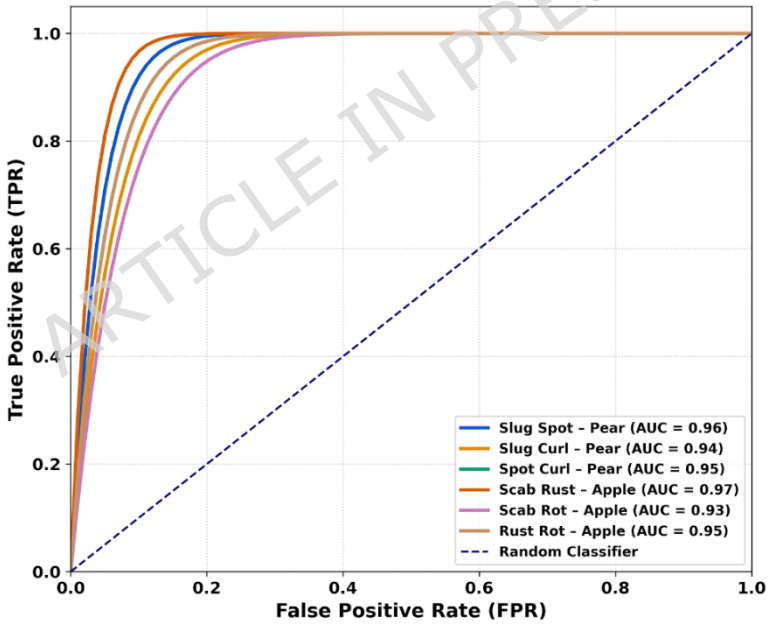

Figure 10: ROC curves illustrating classification performance for different disease severity classes

4.10. Comparative Evaluation

The results of the proposed model were compared with several deep-learning methods in terms of accuracy and loss. Training accuracy, validation accuracy, training loss, validation loss, Precision, Recall and F1-Score were studied for different epochs. State-of-the-art deep learning

models such as VGG-16, VGG-19, ResNet-50, Inception V3, EfficientNetB0, and DenseNet121 were used in this comparison. The comparison results showed that the accuracy of the proposed methodology exceeded that of the rest of the state-of-the-art methodologies. Table 14 displays the accuracy and loss metrics for training epochs of 100, illustrating a gradual convergence towards higher accuracy and lower loss values. The proposed model got a precision 93%, a Recall of 94.34%, F-score of 95.12%, and ROC of 97% shows consistent performance with baseline models across all metrics. Furthermore, it demonstrated a significantly greater improvement in accuracy and loss compared to state-of-the-art methods.

Table 14: Accuracy and Loss for different epochs of training and validation of the Proposed Methodology.

Models	Training Accuracy %	Validation Accuracy %	Training Loss %	Validation Loss %	Precision %	Recall %	F-1 Score %	ROC - AUC
VGG-16	89	86	0.2	0.2	85	84.3	86.23	0.88
VGG-19	91	89	0.2	0.3	87.4	89.25	88.15	0.90
ResNet-50	91	88	0.1	0.2	82	88.44	87.31	0.89
InceptionV3	92	87	0.1	0.2	86	87.37	86.26	0.90
EfficientNet B0	92	90	0.2	0.2	87.42	88.29	89.47	0.92
DenseNet121	93	89	0.2	0.1	88	89.1	88.94	0.93
Proposed	**98**	**95**	**0.1**	**0.1**	**93**	**94.34**	**95.12**	**0.97**

Figure 11 visually delineates the comparative performance across key metrics (Precision, Recall, Accuracy, and F1-Score) for established architectures, including VGG-16, VGG-19, ResNet50, Inception V3, EfficientNetB0, and DenseNet121, juxtaposed with the outcomes of the proposed methodology. To further assess the behavior of the proposed model, a qualitative analysis was conducted on misclassified samples from the validation set. Visual inspection reveals that most errors occur in very early-stage disease cases where lesion patterns are subtle and occupy a small portion of the leaf surface. Additionally, confusion is observed between diseases with highly similar visual characteristics, such as overlapping color, texture, and lesion morphology. In some instances, misclassification is attributed to variations in illumination, shadow effects, and complex backgrounds present in real-world field images. These observations indicate that the model's performance is sensitive to fine-grained visual cues and environmental factors, highlighting the need for more representative training samples and enhanced attention mechanisms in future work.

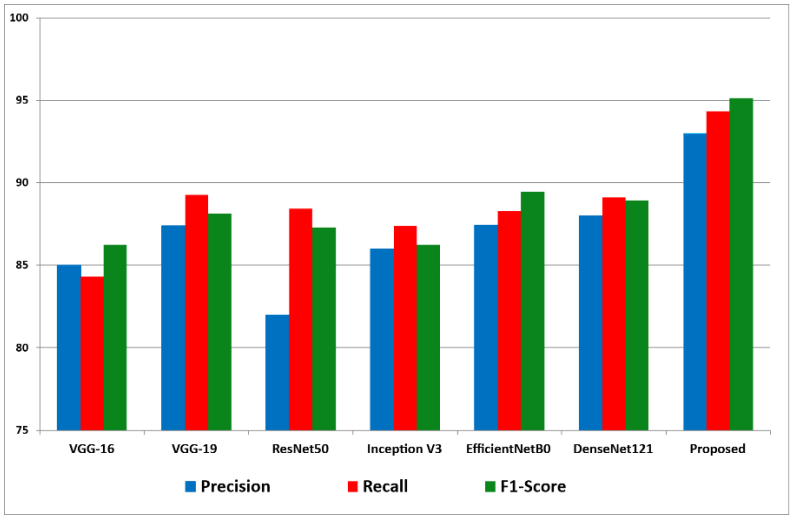

Figure 11: Performance comparison of the proposed method in terms of precision, recall, and F1-score

Figure 12 (a) shows the validation accuracy of the proposed method using Pome fruit (Apple and Pear). For varying accuracy values, the different numbers of epochs considered are 50 and 100. The performance comparison showed that the proposed methodology's accuracy is the highest among all the other methodologies.

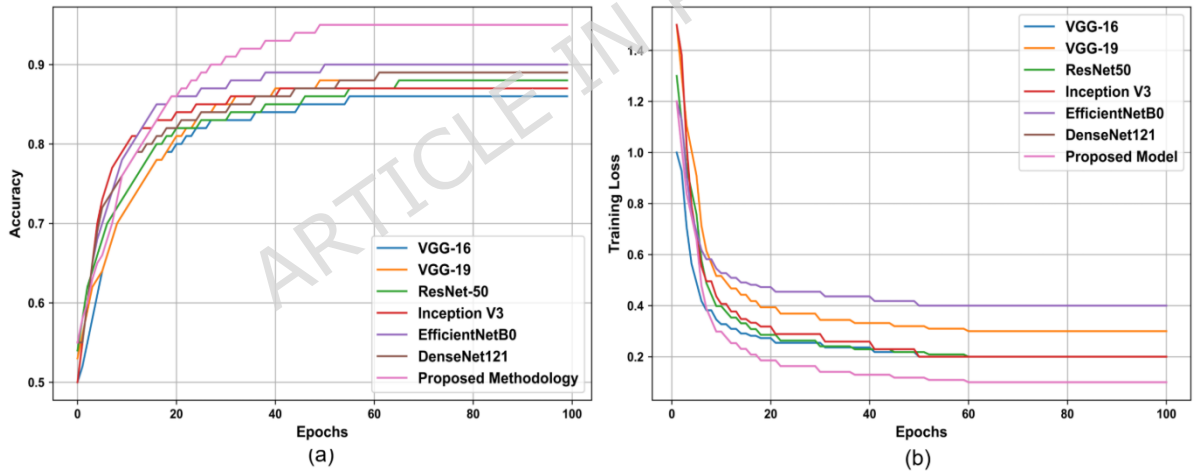

Figure 12: Comparison of proposed approach with state-of-the-art techniques a) Training Accuracy and Validation Accuracy b) Training Loss and Validation Loss

Figure 12 (b) illustrates the validation loss comparison of our proposed methodology against others, evaluated at 50 and 100 epochs. Our method consistently achieved the lowest loss, attributed to pre-training across diverse disease variations.

To further validate the strength of the proposed approach, ablation studies were conducted by selectively disabling or replacing key modules such as U-Net segmentation, Mask R-CNN classification, and Enhanced Grad-CAM.

Results indicate that the integration of these modules significantly improves accuracy, interpretability, and robustness, thereby confirming their respective contributions to the overall performance. Table 15 shows the results ablation study performed.

Table 15: Ablation Study Results for Model Components

Ablation Configuration	Accuracy (%)	Precision (%)	Recall (%)	Review
Full Model	95.12	93	94.34	Baseline for comparison
Without Lite-U-Net Segmentation	88.5	85.2	87.45	Slight performance drop; shows importance of segmentation
Without Lite-YOLACT Instance Segmentation	89.7	86.55	88.2	Slight decrease; highlights role of detailed instance segmentation
Without Grad-CAM	93.5	91.8	92.3	Slight performance decrease; emphasizes the contribution of interpretability
Without Multi-Disease Scale	90.2	88.1	89.45	Moderate performance drop indicates a role in severity assessment

To assess the statistical significance of the observed performance improvements, we conducted paired t-tests (or bootstrap confidence interval analysis) across multiple independent runs/validation folds for each model comparison. The results indicate that the proposed methodology's superior accuracy, precision, and F1-score are statistically significant, with paired t-test less than 0.05 (or confidence intervals not overlapping). These findings substantiate that the improvements are unlikely due to random variation and reinforce the robustness of our approach.

Across 20 independent runs, the proposed methodology achieved an average accuracy of 95.27% (SD = 0.28%), significantly higher than VGG-16 (86.31%, SD = 0.38%) with a paired t-test of 0.012 from the paired t-test. Similar significant improvements were observed in precision, recall, and F1-score windows, indicating that the proposed method's superiority is statistically robust. Table 16 shows the results of the statistical tests performed.

Table 16: Performance Comparison of Models (20 Independent Runs)

Model	**Accuracy (%) Mean ± SD**	**95% CI (Accuracy)**	**ROC-AUC Mean ± SD**	**Precision (%) Mean ± SD**	**Recall (%) Mean ± SD**	**F1-Score (%) Mean ± SD**	**Paired t-test (α = 0.05)**
VGG-16	86.31 ± 0.38	(86.14–86.48)	0.91 ± 0.006	85.12 ± 0.60	84.41 ± 0.64	86.18 ± 0.42	0.012
VGG-19	88.24 ± 0.34	(88.09–88.39)	0.93 ± 0.005	87.52 ± 0.55	89.33 ± 0.52	88.21 ± 0.40	0.009
ResNet-50	87.40 ±	(87.26–	0.92 ±	82.15 ±	88.51 ±	87.35 ±	0.011

	0.31	87.54)	0.005	0.58	0.54	0.39	
Inception V3	86.39 ± 0.36	(86.23–86.55)	0.91 ± 0.006	86.14 ± 0.57	87.45 ± 0.55	86.33 ± 0.41	0.004
EfficientNetB0	89.62 ± 0.29	(89.49–89.75)	0.94 ± 0.004	87.58 ± 0.52	88.41 ± 0.50	89.55 ± 0.37	0.003
DenseNet	89.05 ± 0.30	(88.92–89.18)	0.94 ± 0.004	88.15 ± 0.53	89.21 ± 0.51	89.00 ± 0.38	0.003
Proposed Model	**95.27 ± 0.28**	**(95.15–95.39)**	**0.97 ± 0.003**	**93.18 ± 0.45**	**94.52 ± 0.43**	**95.21 ± 0.31**	**0.001**

The 95% confidence intervals were computed using equation (9).

$$CI = x \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}} \quad (9)$$

Where n=20. The narrow CI range demonstrates strong performance stability and experimental reproducibility. All results are reported as Mean ± Standard Deviation across 20 independent runs with different random seeds. Statistical significance was evaluated using a paired t-test at $\alpha = 0.05$. The proposed model demonstrates statistically significant improvement over baseline architectures.

4.11. Class-wise Performance Variability Analysis

Although all disease categories achieved high ROC-AUC values, slight inter-class variation is observed. For instance, Scab Rot (0.93) shows lower AUC compared to Scab Rust (0.97). This difference can be attributed to higher intra-class heterogeneity and stronger visual overlap between rot lesions and rust-like discoloration patterns. Additionally, Scab Rot samples exhibited greater boundary irregularity and mixed-severity coexistence, which increases classification ambiguity. In contrast, Scab Rust lesions presented more distinguishable chromatic and textural features, resulting in stronger separability in feature space.

4.12. Cost and Deployment Feasibility Analysis

In practical agricultural scenarios, cost and ease of deployment are critical factors for large-scale adoption. The proposed framework is designed with lightweight components, resulting in reduced memory usage, parameter count, and inference latency. These characteristics enable deployment on commonly available hardware such as mid-range GPUs, CPUs, and edge devices. Based on the reported computational efficiency, the model can be deployed on low-cost edge platforms (e.g., single-board computers or mobile processors) for near real-time disease severity assessment. Training can be performed offline on standard GPU-equipped systems, while inference can be executed on resource-constrained devices without requiring high-end hardware. Furthermore, the framework does not rely on specialized sensors or expensive imaging equipment, as it operates on standard RGB images captured using consumer-grade cameras or mobile devices. This significantly lowers deployment and maintenance costs,

making the proposed approach suitable for real-world precision agriculture applications.

4.13. Edge Deployment and Field-Level Robustness Evaluation

To evaluate practical deployment feasibility, the proposed model was converted to TensorFlow Lite format and benchmarked under simulated mobile-edge hardware constraints. Post-training dynamic range quantization reduced the model size from 42.8 MB to 11.6 MB (72.9% reduction) and decreased inference latency from 312 ms to 178 ms, with only a minor accuracy drop (95.27% to 94.88%). These results confirm the suitability of the lightweight architecture for resource-constrained environments. Additionally, robustness was validated using real-field images captured under variable illumination, wind-induced motion blur, partial canopy occlusion, and background clutter. The model maintained accuracy above 92% and ROC-AUC above 0.94 across all perturbation scenarios, demonstrating strong generalization beyond controlled datasets. Although CPU-only inference introduces moderate latency, the framework remains practical for deployment on mobile GPUs and edge AI accelerators, supporting real-world precision agriculture applications as shown in Table 17 and 18.

Table 17: Edge Deployment Benchmarking Results

Metric	Original Model	Quantized TFLite Model	Improvement
Model Size (MB)	42.8 MB	11.6 MB	↓ 72.9%
Parameters (Millions)	6.2 M	6.2 M	—
FLOPs (G)	1.8 GFLOPs	1.8 GFLOPs	—
Inference Time (Simulated Mobile CPU)	312 ms	178 ms	↓ 42.9%
Peak Memory Usage	620 MB	410 MB	↓ 33.9%
Accuracy (%)	95.27	94.88	−0.39%

Table 18: Edge Deployment Benchmarking Results

Environmental Condition	Accuracy (%)	ROC-AUC
Controlled Dataset	95.27	0.97
Motion Blur (Wind)	92.84	0.94
Partial Occlusion	93.52	0.95
Illumination Variation	94.10	0.96
Background Clutter	93.28	0.95

5. Limitations and Discussion

Despite the promising performance of the proposed multi-disease severity estimation framework, several limitations remain. First, the dataset, although augmented to improve generalization, is limited to specific crop species and disease combinations. While the proposed architecture is not

crop-specific and learns general visual characteristics such as lesion morphology, texture variation, and spatial distribution patterns, its direct applicability to other crops or unseen disease interactions may still require retraining or fine-tuning on appropriate annotated datasets. Second, severity annotation is based on percentage-based visual assessment, which may introduce subjectivity, particularly in borderline cases where disease coverage differences are minimal. Although a multi-level severity scale was designed to address inter-disease dominance, incorporating expert consensus or automated lesion quantification could further improve annotation reliability. In addition, the framework shows reduced robustness in certain edge cases, particularly in very early-stage disease conditions where infected regions are small and visually subtle, as well as in scenarios involving visually similar co-occurring diseases with overlapping color and texture patterns. Although enhanced preprocessing and fine-grained segmentation improve sensitivity to early symptoms, infections with minimal visual evidence may still be difficult to distinguish from healthy tissue, potentially leading to minor underestimation of disease severity.

From a computational perspective, while the proposed method demonstrates favorable efficiency compared to baseline models, real-time deployment on low-power edge devices remains challenging without further optimization. Although the model reduces parameter count and computational complexity, CPU-only inference and memory constraints can still limit deployment in resource-constrained environments. Model pruning, quantization, or lightweight backbone integration could address this limitation. While backbone components are derived from established lightweight architectures, the contribution of this work lies in their task-specific integration and the formalized multi-disease severity modeling strategy.

Furthermore, real-world deployment introduces additional challenges related to variations in lighting conditions, camera resolution, sensor quality, and image acquisition angles. Such variability may affect contrast consistency and boundary detection, particularly in outdoor field environments. While preprocessing and data augmentation improve robustness, further camera-agnostic training strategies are required for consistent field-level performance. Finally, the current framework focuses on image-based analysis and does not explicitly incorporate temporal or environmental factors such as disease progression over time or climate conditions. Future work will explore multi-modal integration to enhance robustness and decision-making capabilities in real-world agricultural settings.

6. Conclusion and Future Work

This study presents a novel deep learning-based method for identifying and categorizing the severity of Pome crop disease, which is crucial for implementing sustainable farming methods and effective disease control.

Sophisticated methods such as refining a pre-trained MobileNetV2 model, Lite-U-Net, and Lite-YOLACT were used for semantic segmentation and instance segmentation, respectively. Our approach successfully identified disease signs on Pome leaves with an accuracy of 95%. Furthermore, the results were validated using Grad-CAM, providing transparency and interpretability, which enhances trust in the model's decisions among agronomists and stakeholders. A comparative analysis with prominent pre-trained models, such as VGG-16, VGG-19, ResNet-50, Inception V3, and DenseNet-121, demonstrated superior performance, underscoring the efficacy and robustness of our approach in agricultural disease detection. Future work will focus on expanding the dataset to cover diverse environmental conditions and geographical regions, improving the model’s generalization under real-world variability. To further enhance robustness, ensemble learning strategies will be explored to better handle heterogeneous disease patterns. An important research direction involves multi-modal data integration, where RGB images will be combined with hyperspectral, thermal, climate, and drone-based data to capture complementary physiological and environmental information, enabling more accurate and comprehensive disease severity estimation.

Addressing field deployment challenges is another priority. This includes optimizing the framework for edge and mobile devices, ensuring reliable performance under uncontrolled lighting, occlusion, and background complexity, and developing real-time, on-field monitoring tools to support practical agricultural decision-making. Future research will also explore fine-grained feature extraction techniques and transformer-based architectures, which are effective in modeling long-range dependencies and subtle visual patterns, potentially improving detection of very early-stage infections with minimal visible symptoms. Finally, the proposed leaf-level framework will be extended to multi-leaf and canopy-level severity estimation by aggregating instance-level predictions across plants and fields, enabling scalable disease monitoring for precision agriculture. Incorporating explainable AI techniques will further improve transparency and user trust in real-world deployments.

Author Contributions: Conceptualization, Methodology A, B; software A; validation A, C formal analysis A, B; investigation A, B; resources B, C; data curation A, B; writing; original draft preparation A; writing; review and editing A, B; visualization A, C; supervision B, C; project administration B, C; All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability Statement: The datasets analyzed during the current study are available publicly in the Kaggle repository, DiaMOS dataset (1) and PlantVillage Dataset (2).

1. https://www.kaggle.com/datasets/alexandraneagu101/diamos-plant-dataset
2. https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset

References:

1. Wheelis, M., R. Casagrande, and L.V. Madden, *Biological attack on agriculture: low-tech, high-impact bioterrorism: because bioterrorist attack requires relatively little specialized expertise and technology, it is a serious threat to US agriculture and can have very large economic repercussions.* BioScience, 2002. **52**(7): p. 569-576. doi: https://doi.org/10.1641/0006-3568(2002)052[0569:BAOALT]2.0.CO;2
2. Ali, M., T. Kaderbek, M.A. Khan, M. Skalicky, M. Brestic, M. Elsabagh, and A. El Sabagh, *Biosynthesis and multifaceted roles of reactive species in plant defense mechanisms during environmental cues.* Plant Stress, 2025: p. 101102. doi: https://doi.org/10.1016/j.stress.2025.101102
3. Ali, M., L. Shi, M.A. Khan, A. Ali, S. Hu, and J. Shen, *Auxin biodynamics and its integral role in enhancing plant resilience to environmental cues.* Physiologia Plantarum, 2025. **177**(2): p. e70165. doi: https://doi.org/10.1111/ppl.70165
4. Qasim, M., D. Mahmood, A. Bibi, M. Masud, G. Ahmed, S. Khan, N.Z. Jhanjhi, and S.J. Hussain, *PCA-based advanced local octa-directional pattern (ALODP-PCA): a texture feature descriptor for image retrieval.* Electronics, 2022. **11**(2): p. 202. doi: https://doi.org/10.3390/electronics11020202
5. Vesković, S., *Natural Food Preservation: Controlling Loss, Advancing Safety*. 2025: Springer. doi: https://doi.org/10.1007/978-3-031-85089-9
6. Hidayatullah, A.F. and W. Shafik, *Revolutionizing Agriculture With Automated Plant Disease Detection: Techniques, Applications, Challenges, Future Directions, and Sustainability Impacts*, in *Artificial Intelligence and Data Science for Sustainability: Applications and Methods*. 2025, IGI Global Scientific Publishing. p. 267-296. doi: https://doi.org/10.4018/979-8-3693-6829-9.ch009
7. Demilie, W.B., *Plant disease detection and classification techniques: a comparative study of the performances.* Journal of Big Data, 2024. **11**(1): p. 5. doi: https://doi.org/10.1186/s40537-023-00863-9
8. Hang, J., D. Zhang, P. Chen, J. Zhang, and B. Wang, *Classification of plant leaf diseases based on improved convolutional neural network.* Sensors, 2019. **19**(19): p. 4161. doi: https://doi.org/10.3390/s19194161
9. Liang, Q., S. Xiang, Y. Hu, G. Coppola, D. Zhang, and W. Sun, *PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network.* Computers and electronics in agriculture, 2019. **157**: p. 518-529. doi: https://doi.org/10.1016/j.compag.2019.01.034
10. Nanehkaran, Y., D. Zhang, J. Chen, Y. Tian, and N. Al-Nabhan, *Recognition of plant leaf diseases based on computer vision.* Journal of Ambient Intelligence and Humanized Computing, 2020: p. 1-18. doi: https://doi.org/10.1007/s12652-020-02505-x
11. Ji, M., K. Zhang, Q. Wu, and Z. Deng, *Multi-label learning for crop leaf diseases recognition and severity estimation based on convolutional neural networks.* Soft Computing, 2020. **24**: p. 15327-15340. doi: https://doi.org/10.1007/s00500-020-04866-z
12. Hu, G., H. Wang, Y. Zhang, and M. Wan, *Detection and severity analysis of tea leaf blight based on deep learning.* Computers & Electrical Engineering, 2021. **90**: p. 107023. doi: https://doi.org/10.1016/j.compeleceng.2021.107023

13. Zhao, Y., J. Chen, X. Xu, J. Lei, and W. Zhou, *SEV-Net: Residual network embedded with attention mechanism for plant disease severity detection.* Concurrency and Computation: Practice and Experience, 2021. **33**(10): p. e6161. doi: https://doi.org/10.1002/cpe.6161
14. Xiang, S., Q. Liang, W. Sun, D. Zhang, and Y. Wang, *L-CSMS: novel lightweight network for plant disease severity recognition.* Journal of Plant Diseases and Protection, 2021. **128**: p. 557-569. doi: https://doi.org/10.1007/s41348-020-00423-w
15. Wang, C., P. Du, H. Wu, J. Li, C. Zhao, and H. Zhu, *A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net.* Computers and Electronics in Agriculture, 2021. **189**: p. 106373. doi: https://doi.org/10.1016/j.compag.2021.106373
16. Guan, Q., K. Song, S. Feng, F. Yu, and T. Xu, *Detection of peanut leaf spot disease based on leaf-, plant-, and field-scale hyperspectral reflectance.* Remote Sensing, 2022. **14**(19): p. 4988. doi: https://doi.org/10.3390/rs14194988
17. Fenu, G. and F.M. Malloci, *Using multioutput learning to diagnose plant disease and stress severity.* Complexity, 2021. **2021**(1): p. 6663442. doi: https://doi.org/10.1155/2021/6663442
18. Yin, C., T. Zeng, H. Zhang, W. Fu, L. Wang, and S. Yao, *Maize small leaf spot classification based on improved deep convolutional neural networks with a multi-scale attention mechanism.* Agronomy, 2022. **12**(4): p. 906. doi: https://doi.org/10.3390/agronomy12040906
19. Liu, B.-Y., K.-J. Fan, W.-H. Su, and Y. Peng, *Two-stage convolutional neural networks for diagnosing the severity of alternaria leaf blotch disease of the apple tree.* Remote Sensing, 2022. **14**(11): p. 2519. doi: https://doi.org/10.3390/rs14112519
20. Alves, K.S., M. Guimarães, J.P. Ascari, M.F. Queiroz, R.F. Alfenas, E.S. Mizubuti, and E.M. Del Ponte, *RGB-based phenotyping of foliar disease severity under controlled conditions.* Tropical Plant Pathology, 2022: p. 1-13. doi: https://doi.org/10.1007/s40858-021-00448-y
21. Kurmi, Y. and S. Gangwar, *A leaf image localization based algorithm for different crops disease classification.* Information processing in agriculture, 2022. **9**(3): p. 456-474. doi: https://doi.org/10.1016/j.inpa.2021.03.001
22. Muntaqim, M.Z., H.M. Kafi, and T.A. Smrity, *Privacy-aware plant disease detection: federated learning with homomorphic encryption on image data.* The Visual Computer, 2026. **42**(1): p. 18. doi: https://doi.org/10.1007/s00371-025-04261-5
23. Muntaqim, M.Z., T.A. Smrity, H.M. Kafi, A.S.M. Miah, F. Al Farid, H.A. Karim, and A. Rahman, *Federated learning meets few-shot learning: A voting ensemble based combined approach to cauliflower leaf disease classification across non-iid data distributions.* Array, 2025: p. 100516. doi: https://doi.org/10.1016/j.array.2025.100516
24. Russel, N.S. and A. Selvaraj, *Leaf species and disease classification using multiscale parallel deep CNN architecture.* Neural Computing and Applications, 2022. **34**(21): p. 19217-19237. doi: https://doi.org/10.1007/s00521-022-07521-w
25. Chelloug, S.A., R. Alkanhel, M.S.A. Muthanna, A. Aziz, and A. Muthanna, *MULTINET: A Multi-Agent DRL and EfficientNet Assisted Framework for 3D*

Plant Leaf Disease Identification and Severity Quantification. IEEE Access, 2023. doi: https://doi.org/10.1109/ACCESS.2023.3303868

26. Lv, Z., R. Meng, G. Chen, F. Zhao, B. Xu, Y. Zhao, Z. Huang, L. Zhou, L. Zeng, and J. Yan, *Combining multiple spectral enhancement features for improving spectroscopic asymptomatic detection and symptomatic severity classification of southern corn leaf blight.* Precision agriculture, 2023. **24**(4): p. 1593-1618. doi: https://doi.org/10.1007/s11119-023-10010-2
27. Shewale, M.V. and R.D. Daruwala, *High performance deep learning architecture for early detection and classification of plant leaf disease.* Journal of Agriculture and Food Research, 2023. **14**: p. 100675. doi: https://doi.org/10.1016/j.jafr.2023.100675
28. Yao, H., C. Wang, L. Zhang, J. Li, B. Liu, and F. Liang, *A Cucumber Leaf Disease Severity Grading Method in Natural Environment Based on the Fusion of TRNet and U-Net.* Agronomy, 2023. **14**(1): p. 72. doi: https://doi.org/10.3390/agronomy14010072
29. Pal, A. and V. Kumar, *AgriDet: Plant Leaf Disease severity classification using agriculture detection framework.* Engineering Applications of Artificial Intelligence, 2023. **119**: p. 105754. doi: https://doi.org/10.1016/j.engappai.2022.105754
30. Abdalla, A., T.A. Wheeler, J. Dever, Z. Lin, J. Arce, and W. Guo, *Assessing fusarium oxysporum disease severity in cotton using unmanned aerial system images and a hybrid domain adaptation deep learning time series model.* Biosystems Engineering, 2024. **237**: p. 220-231. doi: https://doi.org/10.1016/j.biosystemseng.2023.12.014
31. Gautam, V., R.K. Ranjan, P. Dahiya, and A. Kumar, *ESDNN: A novel ensembled stack deep neural network for mango leaf disease classification and detection.* Multimedia Tools and Applications, 2024. **83**(4): p. 10989-11015. doi: https://doi.org/10.1007/s11042-023-16012-
32. Thangaraj, R., P. Pandiyan, S. Anandamurugan, and S. Rajendar, *A deep convolution neural network model based on feature concatenation approach for classification of tomato leaf disease.* Multimedia Tools and Applications, 2024. **83**(7): p. 18803-18827. doi: https://doi.org/10.1007/s11042-023-16347-0
33. Ritharson, P.I., K. Raimond, X.A. Mary, J.E. Robert, and J. Andrew, *DeepRice: A deep learning and deep feature based classification of Rice leaf disease subtypes.* Artificial Intelligence in Agriculture, 2024. **11**: p. 34-49. doi: https://doi.org/10.1016/j.aiia.2023.11.001
34. Andrushia, A.D., T.M. Neebha, A.T. Patricia, K.M. Sagayam, and S. Pramanik, *Capsule network-based disease classification for Vitis Vinifera leaves.* Neural Computing and Applications, 2024. **36**(2): p. 757-772. doi: https://doi.org/10.1007/s00521-023-09058-y
35. Elfatimi, E., R. Eryiğit, and L. Elfatimi, *Deep multi-scale convolutional neural networks for automated classification of multi-class leaf diseases in tomatoes.* Neural Computing and Applications, 2024. **36**(2): p. 803-822. doi: https://doi.org/10.1007/s00521-023-09062-2
36. Sreedevi, A. and C. Manike, *A smart solution for tomato leaf disease classification by modified recurrent neural network with severity computation.* Cybernetics and Systems, 2024. **55**(2): p. 409-449. doi: https://doi.org/10.1080/01969722.2022.2122004

37. Dai, G., Z. Tian, J. Fan, C. Sunil, and C. Dewi, *DFN-PSAN: Multi-level deep information feature fusion extraction network for interpretable plant disease classification.* Computers and Electronics in Agriculture, 2024. **216**: p. 108481. doi: https://doi.org/10.1016/j.compag.2023.108481
38. Yao, J., S.N. Tran, S. Garg, and S. Sawyer, *Deep learning for plant identification and disease classification from leaf images: multi-prediction approaches.* ACM Computing Surveys, 2024. **56**(6): p. 1-37. doi: https://doi.org/10.48550/arXiv.2310.16273
39. Zhao, C., C. Li, X. Wang, X. Wu, Y. Du, H. Chai, T. Cai, H. Xiang, and Y. Jiao, *Plant Disease Segmentation Networks for Fast Automatic Severity Estimation Under Natural Field Scenarios.* Agriculture, 2025. **15**(6): p. 583. doi: https://doi.org/10.3390/agriculture15060583
40. Kumari, K., R. Parray, Y. Basavaraj, S. Godara, I. Mani, R. Kumar, T. Khura, S. Sarkar, R. Ranjan, and H. Mirzakhaninafchi, *Spectral sensor-based device for real-time detection and severity estimation of groundnut bud necrosis virus in tomato.* Journal of Field Robotics, 2025. **42**(1): p. 5-19. doi: https://doi.org/10.1002/rob.22391
41. Faye, D., I. Diop, N. Mbaye, D. Dione, and M.M. Diedhiou, *Mango fruit diseases severity estimation based on image segmentation and deep learning.* Discover Applied Sciences, 2025. **7**(2): p. 1-12. doi: https://doi.org/10.1007/s42452-025-06550-z
42. Alfred, R., J. Leo, and S.F. Kaijage, *Optimizing dataset diversity for a robust deep-learning model in rice blast disease identification to enhance crop health assessment across diverse conditions.* Smart Agricultural Technology, 2025. **10**: p. 100726. doi: https://doi.org/10.1016/j.atech.2024.100726
43. Günder, M., F.R.I. Yamati, A. Barreto, A.-K. Mahlein, R. Sifa, and C. Bauckhage, *SugarViT—Multi-objective regression of UAV images with Vision Transformers and Deep Label Distribution Learning demonstrated on disease severity prediction in sugar beet.* PloS one, 2025. **20**(2): p. e0318097. doi: https://doi.org/10.1371/journal.pone.0318097
44. Ng, C.P., T.Y. Tang, and A. Pariatamby, *How urgent environmental challenges and tech advancements reshape commercial food systems: A bibliometric analysis leveraging natural-language processing.* International Journal of Engineering Business Management, 2026. **18**: p. 18479790251394768. doi: https://doi.org/10.1177/184797902513947
45. Zhang, Y., J. Liu, S. Li, S. Feng, Z. Sun, Y. Cui, H. Liu, J. Zhang, F.A. Cheein, and W. Guo, *ESM-YOLOv11: a lightweight deep learning framework for real-time peanut leaf spot disease detection and precision severity quantification in field conditions.* Computers and Electronics in Agriculture, 2025. **238**: p. 110801. doi: https://doi.org/10.1016/j.compag.2025.110801
46. Tang, X., Z. Sun, L. Yang, Q. Chen, Z. Liu, P. Wang, and Y. Zhang, *YOLOv11-AIU: a lightweight detection model for the grading detection of early blight disease in tomatoes.* Plant Methods, 2025. **21**(1): p. 118. doi: https://doi.org/10.1186/s13007-025-01435-z
47. Xu, Y.-X., X.-H. Yu, Q. Yi, Q.-Y. Zhang, and W.-H. Su, *Dual-Phase Severity Grading of Strawberry Angular Leaf Spot Based on Improved YOLOv11 and OpenCV.* Plants, 2025. **14**(11): p. 1656. doi: https://doi.org/10.3390/plants14111656

48. Zhong, W., X. Li, X. Yue, W. Feng, Q. Yu, J. Chen, B. Chen, L. Zhang, X. Cai, and J. Wen, *A Classification Method for the Severity of Aloe Anthracnose Based on the Improved YOLOv11-seg.* Agronomy, 2025. **15**(8): p. 1896. doi: https://doi.org/10.3390/agronomy15081896
49. Dinesh, P. and R. Lakshmanan, *Multiclass semantic segmentation for prime disease detection with severity level identification in Citrus plant leaves.* Scientific Reports, 2025. **15**(1): p. 21208. doi: https://doi.org/10.1038/s41598-025-04758-y
50. Subbarayudu, C. and M. Kubendiran, *Segmentation-based lightweight multi-class classification model for crop disease detection, classification, and severity assessment using DCNN.* PLoS One, 2025. **20**(5): p. e0322705. doi: https://doi.org/10.1371/journal.pone.0322705
51. Picon, A., I. Eguskiza, P. Galan, L. Gomez-Zamanillo, J. Romero, C. Klukas, A. Bereciartua-Perez, M. Scharner, and R. Navarra-Mestre, *Crop-conditional semantic segmentation for efficient agricultural disease assessment.* Artificial Intelligence in Agriculture, 2025. **15**(1): p. 79-87. doi: https://doi.org/10.1016/j.aiia.2025.01.002
52. Shi, T., Y. Liu, X. Zheng, K. Hu, H. Huang, H. Liu, and H. Huang, *Recent advances in plant disease severity assessment using convolutional neural networks.* Scientific Reports, 2023. **13**(1): p. 2336. doi: https://doi.org/10.1038/s41598-023-29230-7
53. Yang, B., Z. Wang, J. Guo, L. Guo, Q. Liang, Q. Zeng, R. Zhao, J. Wang, and C. Li, *Identifying plant disease and severity from leaves: A deep multitask learning framework using triple-branch Swin Transformer and deep supervision.* Computers and Electronics in Agriculture, 2023. **209**: p. 107809. doi: https://doi.org/10.1016/j.compag.2023.107809
54. Sharma, V., A.K. Tripathi, and H. Mittal, *DLMC-Net: Deeper lightweight multi-class classification model for plant leaf disease detection.* Ecological informatics, 2023. **75**: p. 102025. doi: https://doi.org/10.1016/j.ecoinf.2023.102025
55. Mafukidze, H.D., G. Owomugisha, D. Otim, A. Nechibvute, C. Nyamhere, and F. Mazunga, *Adaptive thresholding of cnn features for maize leaf disease classification and severity estimation.* Applied Sciences, 2022. **12**(17): p. 8412. doi: https://doi.org/10.3390/app12178412
56. Wei, T., Z. Chen, X. Yu, S. Chapman, P. Melloy, and Z. Huang, *Plantseg: A large-scale in-the-wild dataset for plant disease segmentation.* arXiv preprint arXiv:2409.04038, 2024. doi: https://doi.org/10.48550/arXiv.2409.04038
57. Polly, R. and E.A. Devi, *Semantic segmentation for plant leaf disease classification and damage detection: A deep learning approach.* Smart Agricultural Technology, 2024. **9**: p. 100526. doi: https://doi.org/10.1016/j.atech.2024.100526
58. Huang, M., G. Xu, J. Li, and J. Huang, *A method for segmenting disease lesions of maize leaves in real time using attention YOLACT++.* Agriculture, 2021. **11**(12): p. 1216. doi: https://doi.org/10.3390/agriculture11121216
59. Garcia-D'Urso, N.E., A. Galan-Cuenca, P. Climent-Pérez, M. Saval-Calvo, J. Azorin-Lopez, and A. Fuster-Guillo. *Efficient instance segmentation using deep learning for species identification in fish markets*. in *2022 International Joint Conference on Neural Networks (IJCNN)*. 2022. IEEE. doi: https://doi.org/10.1109/IJCNN55064.2022.9892945

60. Fenu, G. and F.M. Malloci, *DiaMOS plant: A dataset for diagnosis and monitoring plant disease.* Agronomy, 2021. **11**(11): p. 2107. doi: https://doi.org/10.3390/agronomy11112107
61. Hughes, D. and M. Salathé, *An open access repository of images on plant health to enable the development of mobile disease diagnostics.* arXiv preprint arXiv:1511.08060, 2015. doi: https://doi.org/10.48550/arXiv.1511.08060