



Benchmarking Faster R-CNN Backbones for Explainable Lemon Leaf Disease Detection

Tanawat Palamee¹, Buntueng Yana², and Thanakarn Suangun³

ABSTRACT

Accurate and timely detection of lemon leaf diseases is important for maintaining yield and fruit quality, yet manual scouting is labor intensive and subjective. This study explores a region-based deep learning method for symptom localization by benchmarking Faster R-CNN with various backbone feature extractors, including ResNet50, ResNet101, VGG16, and MobileNetV3, on a dataset of lemon leaf diseases annotated with bounding boxes. To ensure a fair comparison, all models are trained using identical preprocessing, augmentation, and optimization settings. They are evaluated based on detection metrics at multiple Intersection over Union (IoU) thresholds, which include mean Average Precision (mAP) at 0.5 to 0.95, mAP at 0.50, mAP at 0.75, average recall, and mean IoU. Soft Non-Maximum Suppression is also applied during inference to reduce missed detections when lesions appear close together in clusters. Results indicate that Faster R-CNN with a ResNet50 backbone and Soft Non-Maximum Suppression provides the best overall performance, achieving an mAP from 0.5 to 0.95 of 0.6584 and an mAP at 0.75 of 0.8052, while maintaining a strong recall of 0.7326 and high localization quality with a mean IoU of 0.8203. To ensure trustworthy deployment, model explanations are generated using Grad-CAM and LIME. These methods demonstrate that the detector primarily focuses on symptomatic regions instead of background patterns. Rather than proposing a new detector architecture, this work provides a controlled benchmark of backbone selection under a unified Faster R-CNN pipeline. The proposed pipeline presents an effective and interpretable solution for monitoring lemon leaf disease within the evaluated dataset and experimental setting. It also offers practical guidance for selecting backbones in region-based agricultural inspection systems.

Article information:

Keywords: Lemon Leaf Disease, Object Detection, Faster R-CNN, Soft-NMS, Explainable AI

Article history:

Received: January 12, 2026

Revised: March 12, 2026

Accepted: June 12, 2026

Published: June 27, 2026

(Online)

DOI: 10.37936/ecti-cit.2026203.265673

1. INTRODUCTION

Lemon is a high-value horticultural commodity that supports farm income, regional agri-businesses, and export-oriented supply chains. In many producing regions, lemons are cultivated not only for domestic consumption but also for export, making yield stability and fruit quality economically critical [1-3]. However, lemon production continually faces threats from foliar symptoms linked to significant citrus diseases and pest-pathogen complexes. In particular, citrus greening has been widely reported as a severe and costly threat to citrus industries, motivating re-

search into timely monitoring and management to limit productivity losses [4, 5].

Despite the importance of early diagnosis, field scouting and visual inspection remain the most common practices on many farms and in extension settings. These manual processes are labor-intensive, difficult to scale across large orchards, and inherently subjective, leading to variability among inspectors and across environmental conditions [6]. Recent advances in computer vision and deep learning have enabled image-based plant disease analysis with strong performance across many crops, suggesting a practical path toward rapid screening using cameras on

^{1,2,3}The authors are with the Department of Electrical Engineering, School of Engineering, University of Phayao, Phayao, 56000, Thailand, Email: beetanawat7@gmail.com, buntueng.ya@up.ac.th and thanakarn.su@up.ac.th

²Corresponding author: buntueng.ya@up.ac.th

mobile phones, drones, or embedded devices [6, 7]. A significant barrier to adoption is that high-performing deep models are frequently perceived as black boxes. This perception is especially troubling in agricultural contexts, where decisions can impact pesticide application, pruning practices, or quarantine measures, all of which can have direct economic implications. Explainable AI (XAI) has emerged as a critical requirement for the responsible deployment of artificial intelligence systems. It provides human-interpretable evidence regarding the reasoning behind a model's decision to flag a leaf as diseased [8, 9].

From a methodological perspective, modern object detection includes one-stage detectors such as YOLO and SSD, as well as two-stage region-based detectors such as Faster R-CNN. One-stage approaches are typically optimized for speed, whereas two-stage methods often provide stronger localization accuracy, especially when precise region-level reasoning is required [10-12]. In agricultural imagery, symptom regions may be small, irregular, and confounded by background clutter, motivating detectors that explicitly reason over region proposals and pooled region features. Faster R-CNN is a compelling framework for lesion localization because its Region Proposal Network (RPN) and detection head are trained end-to-end and can be paired with different backbone feature extractors while keeping the region-based detection pipeline comparable across architectures [12].

A practical challenge in detecting leaf symptoms is that lesions frequently appear in clusters and may partially overlap. As a result, predictions for lesions that are close to each other can show a significant degree of overlap. Standard Non-Maximum Suppression (NMS) selects the highest-scoring box and removes other boxes whose Intersection over Union (IoU) exceeds a fixed threshold. In order to eliminate duplicate detections of the same lesion, the NMS process entails the selection of the highest-scoring bounding box and the elimination of the others. However, this method may also result in a decrease in recall by omitting accurate detections of adjacent lesions in regions with dense symptoms. To address this problem, we utilize Soft-NMS during inference, which reduces the scores of overlapping boxes rather than removing them. This method enables us to preserve likely detections, even when lesions are situated closely together. It is less reliant on a single threshold and demonstrates greater robustness when multiple true lesions are present in a small area [13].

This study offers a comparative analysis of Faster R-CNN backbones for detecting lemon leaf diseases, utilizing an annotated bounding-box dataset. We benchmark ResNet50, ResNet101, VGG16, and MobileNetV3 feature extractors under a consistent Faster R-CNN pipeline and evaluate performance using standard detection metrics, overlap-based localization measures, and deployment-oriented efficiency

indicators. To improve transparency and practitioner trust, we further integrate XAI methods such as Grad-CAM and LIME to visualize whether model evidence aligns with pathological leaf regions rather than background artifacts [14, 15]. The outcome is both (i) evidence-based guidance on backbone selection for practical deployment and (ii) an explainability-driven workflow to support decision verification in smart farming applications.

To clarify the scope of this paper, we emphasize that the current study serves as a controlled benchmarking study rather than a proposal for a new object detector or a new explainable artificial intelligence (XAI) algorithm. The main contributions of this work are threefold: (1) a fair backbone-level comparison of ResNet50, ResNet101, VGG16, and MobileNetV3 under the same Faster R-CNN detection pipeline, preprocessing strategy, and optimization setting; (2) an analysis of the benefit of Soft-NMS for clustered lesion scenarios in lemon leaf imagery; and (3) an explainability-assisted inspection workflow using Grad-CAM and LIME to verify whether detected instances are associated with disease-relevant regions qualitatively. These contributions aim to offer practical guidance for selecting models in agricultural lesion detection, while also ensuring that the findings are interpreted appropriately.

2. LITERATURE REVIEW

This section reviews prior work that motivates the proposed study and clarifies its research context. We first summarize deep learning approaches for plant leaf disease analysis, emphasizing object detection for symptom localization. We then discuss two-stage region-based detectors, focusing on Faster R-CNN as a standard and modular framework for benchmarking backbone feature extractors. Because agricultural lesions frequently appear in dense or partially overlapping patterns, we also review Soft-NMS as an inference-time improvement over standard NMS. Next, we describe publicly available citrus and lemon leaf datasets and identify the dataset adopted in this work. Finally, we review XAI methods that improve transparency and trust by providing visual and attribution-based evidence for model decisions in agricultural applications.

2.1 Deep learning for leaf disease detection

Deep learning has become the mainstream approach for plant leaf disease analysis because convolutional architectures can learn robust visual cues, such as color changes, texture irregularities, and lesion boundaries, directly from raw images. Early studies focused primarily on image-level classification and reported strong recognition performance across many crops and disease categories when sufficient labeled data were available [6, 7]. Recent agricultural studies have also explored image-level diag-



Fig.1: Representative images and annotations from the Lemon Leaf Diseases dataset.

nosis under limited clean supervision; for example, Janpuangtong and Rattanaopas proposed a multi-round learning framework for durian leaf disease and pest classification that enriches a small expert-labeled dataset with noisy images collected from online sources, demonstrating the potential of weakly supervised data expansion in practical crop disease analysis [16]. As practical needs shifted from recognizing disease presence to identifying where symptoms occur, many studies moved toward localization-aware formulations. Object detection is well-suited to this requirement because it provides both disease identification and spatial localization, supporting symptom verification and downstream decision-making in smart farming.

In agricultural imagery, symptom regions may be small, irregular, and visually confounded by background clutter. Detector performance can therefore be sensitive to scale variation, background complexity, and class imbalance, which may increase false positives or reduce localization accuracy [17]. Prior work suggests several practical strategies to strengthen detection under these conditions. Multi-scale feature aggregation, including feature pyramid designs, can improve sensitivity to small objects [18]. Loss reweighting approaches can mitigate foreground-background imbalance during training [19]. Beyond these general improvements, a direct and widely used lever is the feature extractor. Because region descriptors depend strongly on the backbone representation, swapping backbones offers a controlled way to study the trade-off between accuracy and computational cost without changing the core detector formulation.

Among modern object detectors, one-stage approaches such as YOLO and SSD are designed for high inference speed by predicting class probabilities and bounding boxes directly from dense feature maps in a single forward pass, making them attractive for real-time monitoring and resource-constrained deployment [10, 11]. In contrast, two-stage region-based detectors such as Faster R-CNN [12] explicitly separate proposal generation from final classification and box refinement. This design allows the second stage to focus computation on a smaller set of candidate regions. It often improves localization performance, particularly at stricter overlap criteria and in scenes containing small or clustered targets. This distinction is especially relevant for leaf disease imagery, where symptomatic lesions often occupy only a small

portion of the leaf surface, appear as scattered instances with large intra-class variation in color and texture, and frequently occur in proximity to one another. In such cases, accurate localization is important for symptom verification, severity assessment, and reliable agronomic interpretation [7, 20]. Representative examples from the Lemon Leaf Diseases dataset, along with their corresponding annotations, are presented in Fig. 1. This figure illustrates the visual diversity of symptom appearances, highlights the similarities among certain disease patterns, and emphasizes the crucial need for accurate lesion localization in disease detection.

Faster R-CNN is a widely adopted two-stage framework that integrates an RPN with a region-based detection head in an end-to-end trainable architecture [12]. In Faster R-CNN, a backbone network processes an input image to generate convolutional feature maps. The RPN proposes candidate regions likely to contain objects, and each proposal is mapped to the feature maps and converted into a fixed-length representation using Region of Interest (ROI) pooling [21]. These region features are then used by prediction heads to output class probabilities and refined bounding-box coordinates. This modular design enables Faster R-CNN to be utilized for systematic backbone benchmarking effectively. It allows for comparisons of different feature extractors while ensuring a consistent proposal mechanism and detection head are maintained.

A wide range of backbone families has been adopted in region-based detectors to balance representational power and computational cost, including VGG[22], ResNet[23], Inception[24], DenseNet[25], MobileNet[26], and EfficientNet[27]. These backbones differ in depth, connectivity patterns, and parameter efficiency; their feature quality can substantially influence both detection accuracy (mAP) and localization precision at higher IoU thresholds.

In addition to backbone design, inference-time post-processing is important in dense symptom patterns. NMS can remove neighboring detections too aggressively when lesions overlap or appear close together, which reduces recall. Soft-NMS addresses this limitation by decaying the scores of overlapping boxes rather than discarding them outright, improving robustness in crowded scenes without retraining [13]. This property aligns well with leaf disease imagery, in which multiple symptomatic areas may cluster on a single leaf or appear close to one another.

Table 1: Comparison of representative prior studies and the current work.

Study	Crop	Learning setting	XAI	Main relevance to this work
Ferentinos [7]	Plant diseases across crops	Image-level classification	No	Establishes the effectiveness of deep learning for plant disease recognition.
Rauf <i>et al.</i> [28]	Citrus fruits, leaves, and stems	Dataset and citrus disease resource for machine-learning-based disease analysis	No	Provides a citrus-related benchmark resource for disease analysis.
Siam <i>et al.</i> [29]	Lemon leaves	Dataset paper with image-level nine-class classification baseline using DenseNet-121 and data augmentation	No	Provides lemon-specific dataset context and a baseline classification benchmark relevant to disease analysis.
Gómez-Flores <i>et al.</i> [30]	Orange citrus leaves	Dataset for citrus abnormality detection using image analysis, with binary masks and ROI masks	No	Broadens the citrus abnormality detection context beyond lemon leaves only.
This work	Lemon leaf diseases dataset with bounding-box annotations	Faster R-CNN backbone benchmarking	Grad-CAM, LIME	Controlled empirical comparison of backbone feature extractors under a unified detector pipeline with Soft-NMS and qualitative explanation analysis.

2.2 Citrus and Lemon Leaf datasets

Public datasets play a central role in enabling reproducible evaluations and fair comparisons of plant disease models. For citrus-related research, Rauf *et al.* [28] released a widely used dataset of citrus fruits and leaves covering multiple citrus diseases, which has supported numerous machine learning baselines. Recently, collections centered on lemons have been introduced to more accurately reflect the symptom patterns associated with lemon cultivation. Siam *et al.* [29] provided a comprehensive lemon leaf dataset with multiple disease categories and standardized image resolution, while Gómez-Flores *et al.* [30] released CitrusUAT for citrus leaf abnormality detection, including diseases and nutritional deficiencies.

Although many datasets were initially developed for image-level classification, symptom localization requires object-level annotation to allow detectors to learn features specific to particular regions. In this study, we use the Lemon Leaf Diseases dataset and Annotations [31], which provides bounding boxes for lesion regions. Fig.1 shows representative examples and annotations from the dataset.

2.3 Explainable AI

Trustworthy deployment in agriculture necessitates not only robust detection performance but also interpretable evidence that model decisions are based on pathological regions rather than misleading background cues. As a result, XAI methods are increasingly utilized in conjunction with deep vision models to aid in expert validation and identify failure modes. Grad-CAM provides gradient-based visual explanations by highlighting spatial regions that most influence a prediction [14]. LIME clarifies model behavior by creating a locally interpretable approximation that relies on perturbations of image regions, allowing for the visualization of influ-

ential superpixels [15]. These methods are complementary in practice because they provide different views of explanation, including gradient-based localization and model-agnostic region attribution. In this study, Grad-CAM and LIME are utilized to evaluate whether Faster R-CNN detections target lesion regions. This approach aims to enhance transparency in reporting and boost confidence in the model's predictions for practical applications in smart farming.

To further clarify the novelty and research gap addressed by this study, Table 1 summarizes representative prior works and contrasts them with the current manuscript across crop focus, learning setting, use of XAI, and overall relevance to the present objective. The table indicates that much of the earlier research has focused mainly on image-level disease classification or the development of benchmark datasets for citrus and similar crops. Although these studies provide valuable foundations for agricultural image analysis, they generally do not investigate lesion-level object detection under a unified benchmarking protocol, nor do they examine how alternative backbones feature extractors influence detection performance within the same detector architecture. Additionally, qualitative explainability analysis is seldom included in these earlier studies. The present work focuses on detecting lemon leaf disease at the lesion level using bounding-box annotations. It performs a controlled comparison of Faster R-CNN backbones under consistent hyperparameter settings and applies Soft-NMS to enhance the robustness of post-processing. Additionally, it complements the quantitative evaluation with Grad-CAM and LIME to provide qualitative insights into the behavior of the detector.



Fig.2: Overall experimental workflow of the controlled benchmarking study.

3. METHODOLOGY

This section describes the data preprocessing pipeline and the experimental setup used to benchmark Faster R-CNN backbones for lemon leaf disease detection, including the evaluation protocol and the XAI workflow. To improve clarity for readers, we also summarize the dataset structure in Table 2, present the overall controlled benchmarking workflow in Fig.2, and report the shared hyperparameter settings used across all compared backbones. Fig. 2 offers a clear summary of the experimental procedure, illustrating the sequence from data preprocessing to backbone benchmarking, Soft-NMS inference, cross-validation evaluation, and XAI-based failure analysis, all within a unified pipeline.

3.1 Dataset Description and Summary

We conducted experiments on the Lemon Leaf Diseases dataset [31]. Because the objective of this study is lesion localization, each image is associated with bounding-box annotations that identify symptomatic regions. The dataset used in this work contains 900 images, 900 annotated lesion instances, and 9 disease categories, as summarized in Table 2. Reporting these statistics is crucial because both class distribution and annotation density have a direct impact on the learning difficulty of detectors and the clarity of comparative results. For preprocessing, all images were converted to a uniform format and resized to a standard input scale to ensure consistency during training and inference. Additionally, pixel values were normalized using the mean and standard deviation linked to the pretrained backbone weights utilized in each experiment.

3.2 Data Preprocessing

To improve robustness to real-world acquisition conditions, we applied a consistent augmentation policy across all models. The augmentation set includes geometric transformations that preserve symptom structure, such as horizontal flipping and small-angle rotations, as well as photometric transformations, such as brightness and contrast adjustments. The data preprocessing and augmentation pipeline was consistently maintained across all backbones, ensuring that any variations in performance can be linked to the backbone architecture itself, rather than to discrepancies in input processing. This design is central to the controlled nature of the comparison because it reduces confounding factors and allows the observed performance differences to be interpreted

primarily in terms of feature representation quality rather than data preparation choices.

3.3 Experimental Setup

We adopt a region-based detector following the Faster R-CNN formulation [12]. The input image is processed by a backbone network to generate convolutional feature maps. Following this, a region proposal stage creates candidate regions that may indicate disease symptoms. These region proposals are generated using a RPN. Each proposal is mapped onto the feature maps and converted into a fixed-length representation using ROI pooling. These region features are passed to the detection head, which outputs a disease-class probability distribution and refined bounding-box coordinates.

The detector is optimized using a multi-task objective function defined as:

$$L = L_{cls}(p, p^*) + \lambda p^* L_{reg}(t, t^*) \quad (1)$$

Here, L_{cls} denotes the classification loss for the predicted probability p and the ground-truth label p^* . The regression loss, L_{reg} , measures the deviation between the predicted bounding-box coordinates t and the ground-truth box parameters t^* . The term p^* serves as a binary indicator, assigning a value of 1 to the foreground and 0 to the background. This ensures that the regression penalty is applied only to positive samples. The hyperparameter λ regulates the balance between these two objectives.

To identify the most effective architecture for lemon leaf disease detection, we replaced only the backbone feature extractor while keeping the detection head, proposal strategy, ROI pooling, preprocessing, and evaluation protocol fixed. The benchmarked backbones are ResNet50, ResNet101, VGG16, and MobileNetV3. Each model was initialized with pretraining settings and fine-tuned using the same optimization configuration. This experimental design isolates the influence of feature representation on detection accuracy, sensitivity, localization quality, and deployment efficiency.

All experiments were conducted using PyTorch and executed on an RTX 3080 GPU. To enhance reproducibility, a fixed random seed was employed, and the final model checkpoint was selected based on validation performance. All models were trained under the same optimization settings to ensure a fair comparison: the SGD optimizer, an initial learning rate of 0.001, momentum of 0.9, weight decay of 0.0005, and a total of 100 training epochs with a MultiStepLR

schedule. During inference, Gaussian Soft-NMS with a parameter of 0.5 [13] was applied, and detections with confidence scores below 0.001 were filtered out. The shared experimental settings are summarized in Table 3. Using a common configuration reduces the possibility that one backbone benefits disproportionately from a specially tuned optimization setup.

Table 2: Summary of the Lemon Leaf Diseases dataset.

Item	Value
Total images	900
Annotated instances	900
Number of classes	9
Annotation format	Bounding boxes
Image source	DOI: 10.17632/jkd6t3w635.1

The evaluation of detection performance was conducted using the mAP at an IoU threshold of 0.5, along with averaged mAP across IoU thresholds ranging from 0.5 to 0.95. Precision, recall, and F1-score were calculated based on the matched predictions and ground-truth boxes, adhering to the same IoU matching criteria. The IoU was reported as a measure of localization quality based on overlap.

In addition to overall detection metrics, class-specific diagnostic analyses were conducted using a confusion matrix and ROC curves, which were derived from the final detector outputs following Soft-NMS post-processing. Because each image in the dataset contains a single annotated lesion instance, the highest-confidence predicted bounding box in each image was matched to the ground-truth box using an IoU threshold of 0.5. When the matched prediction met this threshold, its predicted class label was compared with the ground-truth class to fill in the confusion matrix. Predictions that did not meet the IoU criterion were classified as missed detections. For ROC analysis, one-vs-rest curves were created for each class by utilizing the confidence scores from the matched predictions. The corresponding area under the curve (AUC) values were then calculated to summarize the separability of the classes.

Table 3: Shared training and inference settings.

Hyperparameter	Setting
Optimizer	SGD
Initial learning rate	0.001
Momentum	0.9
Weight decay	0.0005
Epochs	100
Learning rate schedule	MultiStepLR
Soft-NMS type	Gaussian
Soft-NMS σ	0.5
Confidence threshold	0.001
Random seed	Fixed
Framework	PyTorch
GPU	RTX3080

To minimize bias resulting from limited data and ensure a fair comparison across models, we applied the same 10-fold cross-validation protocol to all backbones. During each run, the dataset was divided into mutually exclusive folds, and the performance metrics were aggregated across these folds to produce the comparative results reported. This protocol helps reduce dependence on a single split and provides a more reliable estimate of model behavior under the available data regime.

To support trustworthy deployment, we applied XAI methods to interpret detection decisions. Grad-CAM highlights spatial regions that contribute most to a predicted class score by backpropagating gradients to convolutional feature maps [14]. LIME explains predictions by learning a locally interpretable surrogate model based on superpixel perturbations [15]. In this work, explanations are generated for detected instances to verify whether the model attention aligns with pathological leaf regions rather than background artifacts. These methods serve as qualitative post-hoc explanation tools instead of providing formal proof for causal model reasoning. In the context of a region-based detector, they should be interpreted as supportive visual evidence of detector focus, while recognizing that they do not provide a complete theoretical characterization of all internal decision processes.

3.4 Soft-NMS

NMS is a standard post-processing step in object detection pipelines, including region-based detectors. It removes duplicate predictions by retaining the highest-scoring bounding box while suppressing other boxes that overlap beyond a selected IoU threshold [12, 21]. While this greedy strategy is effective for many natural-image settings, it can be overly aggressive in dense scenarios where multiple true instances are close together. In lemon leaf disease imagery, lesions often appear in clusters or partially overlap, so nearby true lesions can produce highly overlapping predictions. Under these conditions, standard NMS may suppress correct neighboring detections and reduce recall, even when the model proposes reasonable bounding boxes.

To alleviate this issue, we employ Soft-NMS at inference time [13]. Instead of discarding overlapping boxes outright, Soft-NMS updates their scores using an overlap-dependent decay, which preserves plausible detections for adjacent lesions while still down-weighting near-duplicate boxes. In this work, we use the Gaussian version of Soft-NMS, which decays the confidence score s_i as a smooth function of IoU:

$$s_i = s_i e^{-\frac{IoU(M, b_i)^2}{\sigma}}, \forall b_i \notin M \quad (2)$$

Here, M denotes the current highest-scoring detection, b_i is a neighboring candidate box, and σ con-

Table 4: Performance comparison of different backbone models.

Backbone	mAP@0.5:0.95	mAP@50	mAP@75	AR	mIoU
ResNet50 + Soft-NMS (Proposed)	0.6584	0.9512	0.8052	0.7326	0.8203
ResNet50	0.6437	0.9330	0.7789	0.7226	0.8203
VGG16	0.6381	0.9667	0.7604	0.7006	0.8162
MobileNetV3	0.6380	0.9435	0.6741	0.7089	0.8182
ResNet101	0.3905	0.8248	0.2608	0.5369	0.7120

trols the decay strength. Boxes with higher overlap with M receive stronger score reductions, whereas boxes with low overlap remain largely unchanged. This behavior is well aligned with clustered symptom patterns because adjacent true lesions are less likely to be eliminated prematurely and can remain in the ranked set for final selection. Importantly, Soft-NMS is a drop-in replacement for standard NMS, does not require retraining, and maintains comparable computational complexity while improving robustness in dense lesion configurations [13].

3.5 Feature Pyramid Network

To improve multi-scale lesion detection, we integrate a Feature Pyramid Network (FPN) with the ResNet-50 backbone [18, 23]. Multi-scale representation is crucial for lemon leaf disease imagery because symptoms can vary significantly, from small specks to larger irregular patches. A single backbone layer often fails to adequately capture all these different sizes. FPN addresses this by constructing a pyramid of feature maps that are both semantically strong and spatially detailed, enabling the detector to localize small lesions while maintaining the ability to recognize larger symptomatic areas [18].

In our implementation, the FPN forms a top-down pathway with lateral connections across the backbone stages, where higher-level features provide rich semantics and are progressively upsampled and fused with lower-level features that preserve fine-grained spatial detail. Specifically, we utilise pyramid levels P2, P3, P4, and P5, which are generated from the corresponding backbone features. These features are aligned using 1×1 lateral convolutions, followed by fusion and 3×3 convolutions to minimise aliasing effects [18]. The resulting pyramid features are then employed by the Region Proposal Network and the detection head, allowing for the proposal and refinement of candidate regions at the most appropriate resolution for each lesion scale [12, 18]. The model is therefore better equipped to detect very small lesions and to avoid missing early indicators of disease that may occupy only a few pixels, while still performing reliably on larger and more diffuse symptomatic areas.

4. RESULTS AND DISCUSSION

Table 4 summarizes the detection and localization performance for the evaluated backbone configura-

tions under a common Faster R-CNN pipeline. The ResNet50 variant with Soft-NMS delivers the most consistent overall results across the primary criteria. It achieves the highest mAP at IoU thresholds of 0.5 to 0.95, with a value of 0.6584. Additionally, it demonstrates the best high-overlap accuracy, reflected in an mAP@0.75 of 0.8052, and boasts the strongest average recall at 0.7326. Relative to the ResNet50 baseline without Soft-NMS, these gains indicate that Soft-NMS preserves plausible neighboring detections in crowded symptom regions and reduces the loss of true positives when multiple lesions occur close together. The improvement is most evident at stricter overlap evaluations, which is where aggressive suppression typically harms performance in clustered lesion patterns.

From a technical standpoint, the results suggest that a moderate-depth backbone offers a favorable balance between representational strength and generalization for the present dataset. ResNet50 seems to offer sufficiently rich features for characterizing lesions while being easier to optimize compared to the deeper ResNet101 configuration, given the current data constraints. In contrast, lighter backbones such as MobileNetV3 remain computationally attractive but show a larger drop under stricter localization criteria, indicating weaker fine-grained boundary representation. We emphasize that this interpretation should be understood as an empirical finding for the current benchmark rather than as a universal rule for all crop-disease detection tasks.

VGG16 achieves the highest mAP@0.50, indicating its effectiveness in producing coarse detections that satisfy a moderate overlap requirement. However, its lower mAP@0.75 suggests that it struggles with box refinement when tighter localization is necessary. MobileNetV3 remains competitive in the mAP@0.5:0.95 range; nonetheless, it experiences a more significant drop at mAP@0.75, which aligns with reduced feature fidelity for accurately placing boundaries in small and low-contrast lesions. ResNet101 consistently underperforms across all metrics. When analyzed alongside its training behavior illustrated in Fig. 6, these results indicate overfitting, as the deeper backbone fits the limited training set well but fails to generalize to validation samples that exhibit different lighting, texture, or appearance.

Qualitative examples in Fig. 3 further illustrate the strengths of the proposed configuration. The detector identifies symptomatic regions on leaves, even



Fig.3: Detection results of complex backgrounds.

in the presence of background clutter, varying lighting conditions, and multiple affected areas on a single leaf. This behavior is important for practical deployment because real field images rarely present clean backgrounds, and reliable localization under such variation supports symptom verification and downstream agronomic interpretation.

To minimize evaluation bias caused by limited data, we utilize ten-fold cross-validation and aggregate the performance results across the various folds. The confusion matrix in Fig.4 indicates strong overall class separability with relatively few errors. The remaining confusions are concentrated among visually similar conditions, which is expected when diseases share characteristics such as mottling, discoloration, and subtle texture changes. To qualitatively examine these errors, Fig. 5 presents representative examples of misclassifications. It can be seen that the detector generally identifies the leaf region accurately, while the classification branch assigns an incorrect class label. This indicates that the primary source of error is class-level ambiguity rather than a failure in localization. Such confusion may stem from overlapping visual traits, including similar color patterns, textures, and lesion distributions. These findings point to practical directions for improvement, including expanding the dataset with more diverse field conditions and refining bounding-box annotations for ambiguous lesion boundaries.

Fig. 6 compares the optimization dynamics of different backbone architectures throughout the training schedule. The MobileNetV3-based model begins with a higher loss and shows a slower convergence rate, whereas the ResNet50-based variants and VGG16 achieve a lower loss more quickly and maintain stability. The proposed ResNet50 with Soft-NMS exhibits stable convergence, which aligns with its stronger high-overlap localization performance and higher re-

Dry Leaf	180	0	0	0	0	0	0	0	0
Spider Mites	0	150	10	0	0	0	0	0	20
Curl Virus	0	0	180	0	0	0	0	0	0
Healthy Leaf	0	0	0	180	0	0	0	0	0
Anthracnose	0	0	0	0	180	0	0	0	0
Sooty Mould	0	0	0	0	0	160	0	0	20
Deficiency Leaf	0	0	0	0	0	0	170	10	0
Citrus Canker	0	0	0	0	0	0	0	180	0
Bacterial Blight	0	0	0	0	0	0	0	0	180

Fig.4: Confusion matrix of the proposed model.

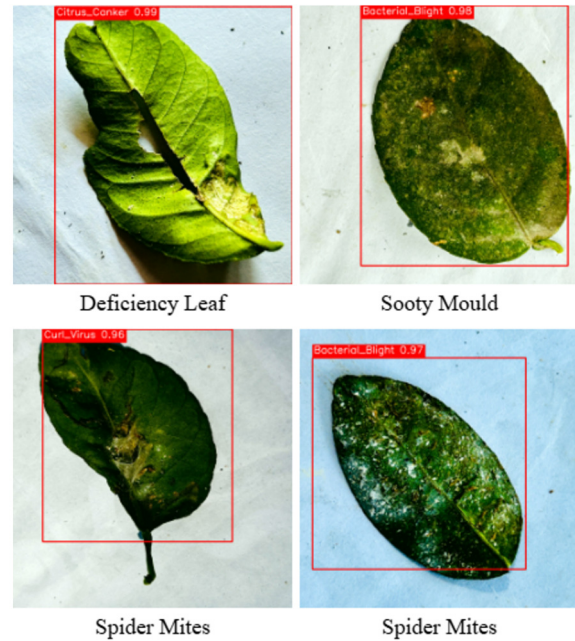


Fig.5: Examples of misclassified detections produced by the proposed model.

call reported in Table 4. The mismatch between low training loss and weak validation metrics for ResNet101 supports the observation that increased depth does not necessarily translate to better detection performance with limited training data.

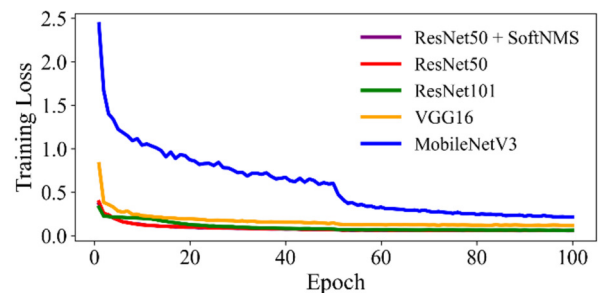


Fig.6: Training loss across backbone configurations.

Fig. 7 displays receiver operating characteristic curves that summarize class separability across different decision thresholds. The consistently high curves suggest that the detector retains strong discrimination capability among categories, which facilitates reliable screening in practical monitoring environments where operating points can be modified to emphasize either recall or precision.

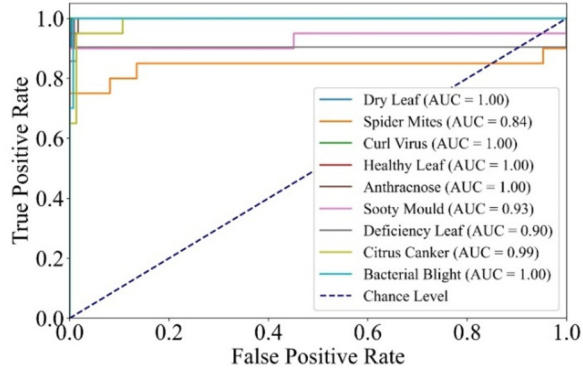


Fig.7: ROC curves across classes.

Finally, the interpretability results in Fig.8 provide representative explanation maps across disease categories. The highlighted regions typically correspond to visible lesion areas rather than background structures, suggesting that the predictions are mainly influenced by cues relevant to the disease. This evidence supports practitioner trust and provides a diagnostic tool for analyzing failure cases, especially for borderline predictions under challenging illumination or when the symptoms are visually subtle. At the same time, these maps should be viewed as qualitative supportive evidence rather than conclusive proof of detector reasoning. Their primary role in this study is to verify whether the detector appears to attend symptom-relevant image regions and to assist error inspection in difficult cases.

5. CONCLUSION

This study assesses various backbone feature extractors for Faster R-CNN using a lemon leaf disease dataset annotated with bounding boxes and follows a unified training and evaluation protocol. Among the tested configurations, ResNet50 with Soft-NMS provides the most reliable overall performance, achieving an $\text{mAP}@0.5:0.95$ of 0.6584, an $\text{mAP}@0.75$ of 0.8052, and the highest average recall of 0.7326. The results show a strong balance between detection accuracy and precise localization, with clear benefits when lesions are clustered or partially overlapping.

To support practical agricultural inspection, the study incorporates explainable AI methods that attach visual evidence to detector outputs. The explanation maps typically emphasize lesion regions and symptom textures rather than background patterns, improving transparency and helping users vali-

date detections. This interpretability component also makes it easier to inspect borderline cases and understand common failure modes.

Rather than introducing a new detector architecture, the present work contributes a controlled empirical benchmark for backbone selection under a fixed Faster R-CNN pipeline. The main empirical finding is that, for the evaluated dataset and protocol, a moderate-depth backbone such as ResNet50 offers a stronger balance between feature quality, localization precision, and generalization than either a lighter or deeper alternative. The conclusion drawn should be understood in the context of the current dataset and should not be applied to all crop-disease detection scenarios without further validation.

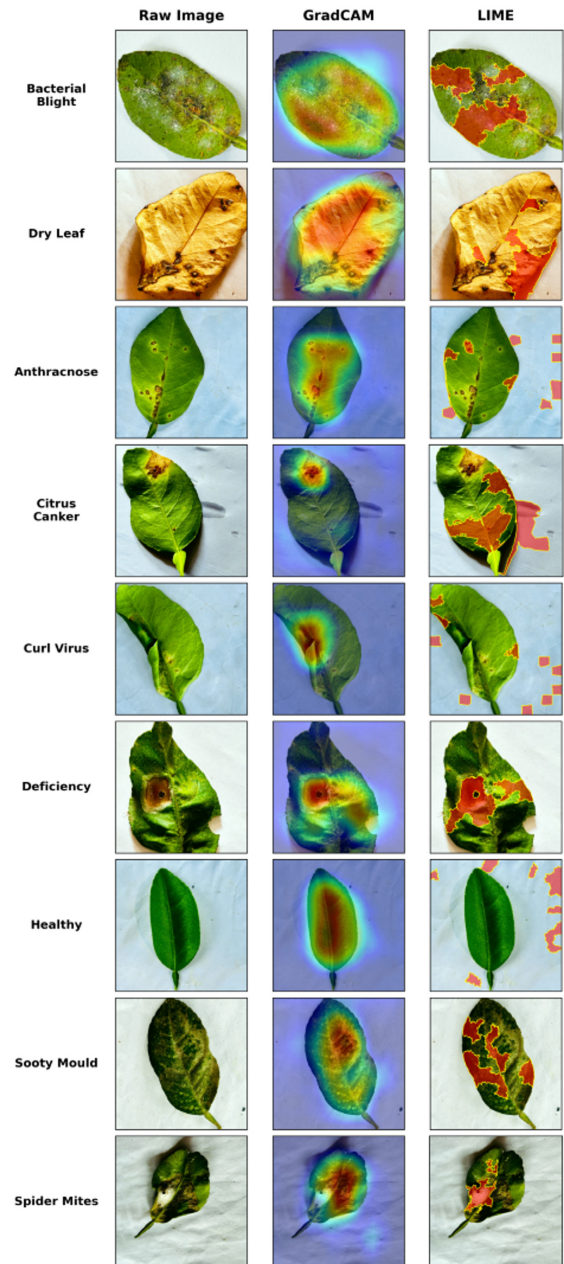


Fig.8: XAI maps across disease categories.

Several limitations remain. The dataset remains limited in diversity compared to real-world conditions. Additionally, small-lesion bounding boxes may encompass surrounding healthy tissue, which can compromise precise localization and heighten confusion among similar diseases. Future work will expand field data across varying environments, strengthen annotation quality, and evaluate efficient deployment on resource-limited devices while exploring improved multi-scale learning and class-imbalance handling to reduce errors further. Additional validation on other citrus datasets, orchards, and imaging conditions will also be necessary before making claims regarding robustness and real-world deployment.

AUTHOR CONTRIBUTIONS

Conceptualization and methodology: Tanawat Palamee and Buntueng Yana; Software: Tanawat Palamee; Data curation: Tanawat Palamee and Buntueng Yana; Writing-original draft preparation: Tanawat Palamee and Buntueng Yana; Writing-review and editing: Buntueng Yana and Thanakarn Suangun; Supervision: Buntueng Yana.

References

- [1] M. I. Cabot, J. Lado and N. Sanjuán, “Multi-season environmental life cycle assessment of lemons: A case study in south Uruguay,” *Journal of Environmental Management*, vol. 326, p. 116719, 2023.
- [2] R. Ciriminna, G. Angellotti, R. Luque and M. Pagliaro, “The citrus economy in Sicily in the early bioeconomy era: a case study for bioeconomy practitioners,” *Biofuels, Bioproducts & Biorefining*, vol. 18, no. 2, pp. 356–364, 2024.
- [3] E. Schimmenti, V. Borsellino and A. Galati, “Growth of citrus production among the Euro-Mediterranean countries: political implications and empirical findings,” *Spanish Journal of Agricultural Research*, vol. 11, no. 3, pp. 561–577, 2013.
- [4] S. Alvarez, E. Rohrig, D. Solis and M. Thomas, “Citrus Greening Disease (Huanglongbing) in Florida: Economic Impact, Management and the Potential for Biological Control,” *Agricultural Research*, vol. 5, no. 2, pp. 109–118, 2016.
- [5] C. Monzó and P. A. Stansly, “Economic value of conservation biological control for management of the Asian citrus psyllid, vector of citrus Huanglongbing disease,” *Pest Management Science*, vol. 76, no. 5, pp. 1691–1698, 2020.
- [6] A. Abade, P. A. Ferreira and F. de Barros Vidal, “Plant diseases recognition on images using convolutional neural networks: A systematic review,” *Computers and Electronics in Agriculture*, vol. 185, p. 106125, 2021.
- [7] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.
- [8] A. Adadi and M. Berrada, “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),” in *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
- [9] A. Barredo Arrieta *et al.*, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, 2020.
- [10] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 779–788, 2016.
- [11] W. Liu *et al.*, “SSD: Single Shot MultiBox Detector,” in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, pp. 21–37, 2016.
- [12] S. Ren, K. He, R. Girshick and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28, pp. 91–99, 2015.
- [13] N. Bodla, B. Singh, R. Chellappa and L. S. Davis, “Soft-NMS — Improving Object Detection with One Line of Code,” *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 5562–5570, 2017.
- [14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 618–626, 2017.
- [15] M. T. Ribeiro, S. Singh and C. Guestrin, ““Why should I trust you?”: Explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016.
- [16] S. Janpuangtong and K. Rattanaopas, “Leveraging Multi-Round Learning and Noisy Labeled Images from Online Sources for Durian Leaf Disease and Pest Classification,” *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol. 19, no. 1, pp. 108–120, 2025.
- [17] S. Dananjayan, Y. Tang, J. Zhuang, C. Hou and S. Luo, “Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images,” *Computers and Electronics in Agriculture*, vol. 193, p. 106658, Feb. 2022.

- [18] T. -Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 936-944, 2017.
- [19] T. -Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 2999-3007, 2017.
- [20] S. P. Mohanty, D. P. Hughes and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, p. 1419, 2016.
- [21] R. Girshick, "Fast R-CNN," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440-1448, 2015.
- [22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [23] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, 2016.
- [24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 2818-2826, 2016.
- [25] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 2261-2269, 2017.
- [26] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [27] P. B. Kolhe, D. S. Ramesh and N. Agarwal, "EnhanceNet: Rethinking Model Scaling for Convolutional Neural Network," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pp. 6105-6114, 2019.
- [28] H. T. Rauf, B. Saleem, M. I. Lali, M. Khan, M. Sharif and S. Bukhari, "A Citrus Fruits and Leaves Dataset for Detection and Classification of Citrus Diseases through Machine Learning," *Data in Brief*, vol. 26, p. 104340, 2019.
- [29] A. K. M. F. K. Siam *et al.*, "A comprehensive image dataset for the identification of lemon leaf diseases and computer vision applications," *Data in Brief*, vol. 58, p. 111244, 2025.
- [30] W. Gómez-Flores, J. J. Garza-Saldaña, and S. E. Varela-Fuentes, "CitrusUAT: A dataset of orange Citrus sinensis leaves for abnormality detection using image analysis techniques," *Data in Brief*, vol. 52, p. 109908, 2024.
- [31] B. Doh, "Lemon Leaf Diseases Dataset and Annotations," *Mendeley Data*, v1, 2025. doi: 10.17632/jkd6t3w635.1.



Tanawat Palamee received his B.E. degree in Electrical Engineering from the University of Phayao, Thailand. His research interests span multiple domains, focusing on the intersection of intelligent systems and real-world problem-solving. His research areas include developing algorithms for weed detection, applying machine learning techniques to applications, and exploring SLAM algorithms. He is currently preparing to pursue a master's degree in robotics and machine learning to further his expertise in autonomous systems.



Buntueng Yana is a Lecturer in the Department of Electrical Engineering, Faculty of Engineering, University of Phayao, Thailand, where he has served since 2009. He received his PhD in Information Science and Technology from Osaka University, Japan (2019), and his M.Eng. and B.Eng. in Electrical Engineering from Chiang Mai University, Thailand (2009 and 2006). His expertise includes electronic circuit and embedded-system design, digital communications, wireless networking, artificial intelligence, data analytics, and satellite image processing. His research focuses on robotics, machine learning, and automatic control systems.



Thanakarn Suangun received B. Eng., M.Sc. in Communication Engineering and a Ph.D. degree in Electrical Engineering from King Mongkut's University of Technology North Bangkok (KMUTNB), Thailand, in 2006, 2009, and 2020, respectively. Since 2012, he has been a lecturer at the Department of Electrical Engineering, School of Engineering, University of Phayao, Thailand. His research interests include RF/microwave, multiband small antennas for communication applications, and automation.