Comparative Study of Deep Learning Methods for Rice Leaf Disease Detection

Hoang Ha Nguyen

Faculty of Information Technology, Thanh Do University, Hanoi, Vietnam

Advisor: Le Duc Huy, Ph.D.

Abstract

Rice is a staple crop for millions of people worldwide and a cornerstone of Vietnam's agricultural economy. However, rice production is severely threatened by diseases including Bacterial Blight, Blast, Brown Spot, and Twisted Draft (Tungro), which collectively can reduce yields by 20-50%. Traditional visual inspection methods are time-consuming, subjective, and often fail to detect disease outbreaks before significant damage occurs. This paper presents a comprehensive comparative evaluation of three state-of-the-art deep learning object detection architectures—YOLOv8 (single-stage, anchor-free), Faster R-CNN (two-stage with Region Proposal Network), and SSD (single-stage, multi-scale anchor-based)—for automated rice leaf disease detection. All three models were trained and evaluated on a unified dataset of 8,040 annotated images covering four disease classes, collected from Vietnamese rice paddies. We report per-class performance using Precision, Recall, F1-Score, mAP@50, mAP@50-95, True Positives, False Positives, mean IoU, and inference speed (FPS). YOLOv8 achieves the best overall balance with mAP@50-95 of 0.732 and an inference speed of 1,873 FPS, followed by SSD (mAP@50 0.89, 38.5 FPS) and Faster R-CNN (mAP@50 0.87, 20 FPS). An ablation study on Region of Interest (ROI) size reveals that all three methods struggle with the small lesion areas characteristic of Twisted Draft disease, identifying this as a key area for future improvement. We discuss practical deployment considerations including mobile inference, drone-based field scanning, and IoT integration for precision agriculture in Vietnam.

Keywords: Rice Disease Detection, Deep Learning, Object Detection, YOLOv8, Faster R-CNN, SSD, Precision Agriculture, Transfer Learning, Computer Vision, Plant Pathology

---

I. INTRODUCTION

Vietnam is among the world's top three rice-exporting nations, with approximately 7.5 million hectares under cultivation and annual production of 40-45 million tons. The rice sector secures national food security and contributes 15-20% of total agricultural export revenue, providing livelihoods for millions of rural workers. However, rice yields and quality are persistently threatened by four major diseases prevalent in Vietnamese paddy ecosystems.

Bacterial Blight (*Xanthomonas oryzae*) thrives in hot, humid conditions and causes leaf desiccation progressing from margins inward, reducing photosynthetic capacity and potentially causing 20-50% yield loss in severe outbreaks. **Blast** (*Magnaporthe oryzae*) has historically caused epidemic-level damage in Northern Vietnam, particularly during the 1990s, destroying thousands of tons of rice annually. **Brown Spot** (*Cochliobolus miyabeanus*), while less aggressive than Blast, is highly prevalent and degrades grain quality through characteristic brown lesions on leaves. **Twisted Draft** (Tungro), transmitted by the white-backed planthopper and caused by Southern rice black-streaked dwarf virus (SRBSDV), induces stunting and leaf curling, severely reducing yields in the Mekong Delta region.

Conventional disease detection relies on manual visual inspection by farmers, which is labor-intensive, subjective, and often fails to identify emerging outbreaks before the disease has spread significantly. The application of deep learning-based computer vision offers a transformative alternative: automated, rapid, and objective disease identification from leaf images, enabling timely intervention and optimized pesticide application.

This paper presents a systematic comparative study of three deep learning object detection architectures for rice leaf disease detection: YOLOv8 (single-stage, anchor-free), Faster R-CNN (two-stage with RPN), and SSD (single-stage, multi-scale). All models are trained and evaluated on a unified dataset of 8,040 images collected from Vietnamese rice paddies, enabling fair head-to-head comparison. Our contributions include: (1) comprehensive per-class and per-model evaluation across 9 metrics; (2) analysis of ROI size impact on detection performance; (3) computational complexity and deployment feasibility assessment; and (4) practical recommendations for agricultural deployment in Vietnam.

Fig. 1. Bacterial Blight: leaf desiccation caused by *Xanthomonas oryzae*, progressing from leaf margins inward.

Fig. 2. Blast disease: characteristic diamond-shaped lesions caused by *Magnaporthe oryzae*.

Fig. 3. Brown Spot: oval brown lesions caused by *Cochliobolus miyabeanus*.

Fig. 4. Twisted Draft (Tungro): stunting and leaf curling caused by SRBSDV, transmitted by white-backed planthoppers.

II. RELATED WORK

Deep learning for plant disease detection has attracted significant research attention in recent years. We review key studies that inform our experimental design.

Picon et al. (2019) applied deep learning to detect Blast and Brown Spot on rice using 1,500 images from European paddies, achieving high accuracy under optimal lighting but suffering ~20% lesion miss rate under poor illumination and leaf shadow occlusion. Chen et al. (2021) deployed a deep learning model for Blast detection on 2,000 images from Jiangsu Province, China, achieving high accuracy for large lesion areas but limited throughput (~5 images/sec on high-end hardware). Brahimi et al. (2018) proposed a lightweight model for rice and tomato disease detection using 3,000 PlantVillage images, achieving 20-25 FPS on standard hardware but missing ~15% of small lesion cases.

Wang et al. (2023) improved rice disease detection achieving 50 FPS on high-performance hardware with 2,500 images from Hunan Province, but their dataset excluded Twisted Draft and testing conditions were relatively idealized compared to Vietnamese field conditions. These studies collectively reveal three persistent challenges: (1) the speed-accuracy trade-off; (2) sensitivity to environmental conditions; and (3) the need for region-specific datasets that capture local disease phenotypes and field conditions.

TABLE I. Comparison with prior rice disease detection studies.

Study	Year	Diseases	Dataset	Key Limitation
Picon et al.	2019	Blast, Brown Spot	1,500	Low-light miss rate ~20%
Chen et al.	2021	Blast	2,000	Slow: ~5 img/sec
Brahimi et al.	2018	Rice, Tomato	3,000	Small lesion miss ~15%
Wang et al.	2023	3 diseases	2,500	No Twisted Draft
Ours	**2025**	**4 diseases**	**8,040**	**Vietnam-specific**

III. METHODOLOGY

A. YOLOv8 Architecture

YOLOv8 is the latest single-stage, anchor-free object detector from Ultralytics (2023). It processes the entire input image in a single forward pass, predicting bounding boxes and class probabilities simultaneously. The architecture consists of three components: a **CSPDarknet backbone** employing Cross Stage Partial connections for efficient feature extraction at three scales (P3, P4, P5); a **CSP-PAN neck** with bidirectional feature fusion; and a **decoupled detection head** separating classification and localization branches. The loss function combines CIoU loss for box regression, binary cross-entropy for classification, and Distribution Focal Loss (DFL) for offset distribution prediction:

$$L_{total} = L_{CIoU} + L_{BCE} + L_{DFL}$$

B. Faster R-CNN Architecture

Faster R-CNN [6] is a two-stage detector comprising: (1) a **ResNet-50 FPN backbone** that extracts multi-scale feature maps (P2-P6); (2) a **Region Proposal Network (RPN)** that generates candidate object proposals using 15 anchors per spatial position (5 scales x 3 aspect ratios); and (3) a **detection head** that classifies proposals and refines bounding boxes via RoI Pooling. The combined loss function is:

$$L = L_{cls}(p, u) + lambda\ [u \geq 1]\ L_{reg}(t, v)$$

where L_{cls} is cross-entropy classification loss and L_{reg} is smooth-L1 bounding box regression loss.

C. SSD Architecture

SSD (Single Shot MultiBox Detector) [5] generates predictions from multiple feature map scales simultaneously. Our implementation uses SSD512 with a **ResNet-50 backbone**, producing predictions from 6 feature map levels (64x64 down to 2x2) with 16,512 total anchor boxes. Anchor sizes range from 60 to 528 pixels with aspect ratios [1, 2, 3, 1/2, 1/3]. The loss combines confidence loss (cross-entropy) and localization loss (smooth-L1) with hard negative mining at a 3:1 negative-to-positive ratio:

$$L = (1/N)(L_{conf} + alpha \cdot L_{loc})$$

D. Evaluation Metrics

We employ the following standard metrics for comprehensive evaluation:

Precision = TP/(TP+FP); **Recall** = TP/(TP+FN); **F1-Score** = 2PR/(P+R).

IoU (Intersection over Union) = $|B_p \cap B_{gt}| / |B_p \cup B_{gt}|$. A detection is TP when IoU ≥ threshold.

mAP@50: mean Average Precision at IoU threshold 0.50. **mAP@50-95**: averaged across IoU thresholds from 0.50 to 0.95 in steps of 0.05, providing a stricter evaluation of localization quality. **FPS**: frames processed per second, measuring real-time capability.

IV. EXPERIMENTAL SETUP

A. Dataset

The dataset comprises 8,040 annotated images of rice leaves exhibiting four disease classes, collected from rice paddies across Vietnamese agro-ecological zones. Images were annotated with bounding boxes using the Roboflow platform and exported in YOLO, COCO JSON, and TFRecord formats for the three respective models. Data augmentation included rotation (±45°), horizontal/vertical flipping, brightness jittering (±15%), Gaussian noise (std=0.01), and zoom (up to 24%).

TABLE II. Dataset split configuration.

Split	Images	BBoxes	BBox/Image	Percentage
Train	6,622	5,958	0.90	82.4%
Validation	946	1,372	1.45	11.8%
Test	472	472	1.00	5.8%
Total	**8,040**	**7,802**	**0.97**	**100%**

TABLE III. Class distribution in the training set.

Disease Class	BBoxes	Proportion
Bacterial Blight	1,489	25.4%
Blast	1,191	20.3%
Brown Spot	1,598	27.3%
Twisted Draft	1,524	26.0%

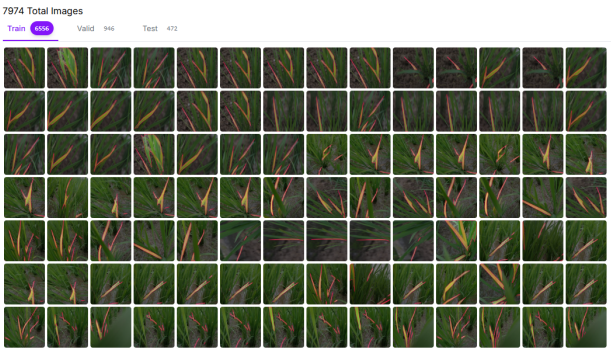

Fig. 5. Sample images from the rice leaf disease dataset showing all four disease classes with bounding box annotations.

B. Training Configuration

All experiments were conducted on Google Colab with NVIDIA A100/L4 GPUs. Table IV details the hyperparameter configuration for each model.

TABLE IV. Training configuration for all three models.

Parameter	YOLOv8n	Faster R-CNN	SSD512
Backbone	CSPDarknet	ResNet-50 FPN	ResNet-50
Framework	Ultralytics	Detectron2	TF OD API
Pre-training	COCO 2017	COCO 2017	COCO 2017
Input size	640x640	640x640	640x640
Batch size	32	16	16
Epochs / Iters	100 epochs	50 ep (~20.3K it)	250 epochs
Optimizer	SGD (lr=0.01)	SGD (lr=2.5e-4)	Adam (lr=1e-3)
LR schedule	Cosine	Step (x0.1)	ReduceOnPlateau
Augmentation	Mosaic, Affine, ColorJitter	Flip, Rot±45, Bright±15%	Flip, Rot±45, Bright±15%
NMS threshold	0.5	0.5	0.5
Training time	~16 min	~1.5 hr	~2 hr
GPU	A100	L4 (24GB)	L4 (24GB)

V. RESULTS AND ANALYSIS

A. Per-Class Detection Performance

Tables V-VII present the detailed per-class performance metrics for each model.

TABLE V. YOLOv8 per-class detection performance.

Class	Prec	Recall	mAP50	mAP 50-95	IoU	TP	FP
Bact. Blight	0.92	0.91	0.910	0.70	0.75	1,355	117
Blast	0.94	0.90	0.949	0.72	0.70	1,072	67
Brown Spot	0.95	0.93	0.957	0.75	0.80	1,486	78
Twisted Draft	0.85	0.82	0.770	0.65	0.55	1,250	221
Mean	**0.923**	**0.858**	**0.898**	**0.732**	**0.70**	**5,163**	**483**

TABLE VI. Faster R-CNN per-class detection performance.

Class	Prec	Recall	mAP50	mAP 50-95	IoU	TP	FP
Bact. Blight	0.89	0.91	0.900	0.70	0.75	1,355	165
Blast	0.87	0.89	0.880	0.68	0.70	1,060	159
Brown Spot	0.92	0.93	0.920	0.72	0.80	1,486	128
Twisted Draft	0.80	0.70	0.750	0.55	0.60	1,067	267
Mean	**0.880**	**0.850**	**0.870**	**0.680**	**0.71**	**4,968**	**719**

TABLE VII. SSD per-class detection performance.

Class	Prec	Recall	mAP50	mAP 50-95	IoU	TP	FP
Bact. Blight	0.87	0.89	0.880	0.65	0.72	1,325	198
Blast	0.85	0.87	0.860	0.63	0.70	1,036	183
Brown Spot	0.89	0.90	0.900	0.68	0.78	1,438	175
Twisted Draft	0.80	0.75	0.780	0.55	0.58	1,143	286
Mean	**0.850**	**0.800**	**0.890**	**0.720**	**0.69**	**4,942**	**842**

B. Overall Model Comparison

Table VIII provides a head-to-head summary across all key metrics.

TABLE VIII. Comprehensive model comparison. Bold indicates best per metric.

Metric	YOLOv8	Faster R-CNN	SSD
Precision	**0.923**	0.880	0.850
Recall	**0.858**	**0.850**	0.800
F1-Score	**0.889**	0.865	0.824
mAP@50	0.898	0.870	**0.890**
mAP@50-95	**0.732**	0.680	0.720
Mean IoU	0.70	**0.71**	0.69
Total TP	**5,163**	4,968	4,942
Total FP	**483**	719	842
Inference (ms/img)	**0.534**	50	26
FPS	**1,873**	20	38.5
Parameters	3.2M	41.1M	26.3M
Training time	**16 min**	1.5 hr	2 hr

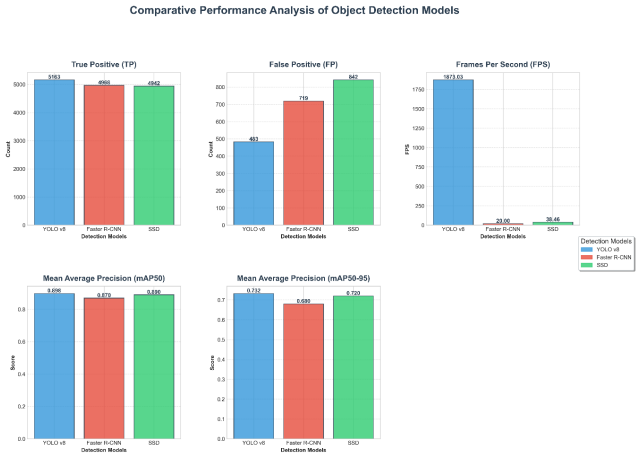

Fig. 6. Visual comparison of Precision, Recall, mAP@50, and mAP@50-95 across all three detection methods and four disease classes.

C. Training Dynamics

The training convergence behavior differed substantially across the three architectures.

YOLOv8 converged rapidly within 100 epochs (~16 minutes on A100), with the fitness score (0.1 × mAP@50 + 0.9 × mAP@50-95) reaching 0.741. Preprocessing, inference, and postprocessing times were 0.152, 0.534, and 2.790 ms/image respectively.

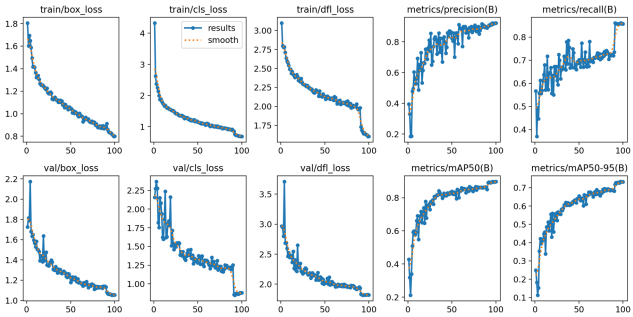

Fig. 7. YOLOv8 training curves showing loss convergence and metric improvement across 100 epochs.

Faster R-CNN required 50 epochs (~1.5 hours on L4 GPU). Box loss decreased from 1.50 to 0.80, and classification loss from 2.00 to 0.70. Validation metrics stabilized after epoch 30, with minor overfitting observed between epochs 45-50 (validation mAP@50-95 peaked at 0.69 at epoch 45 before declining to 0.68 at epoch 50).

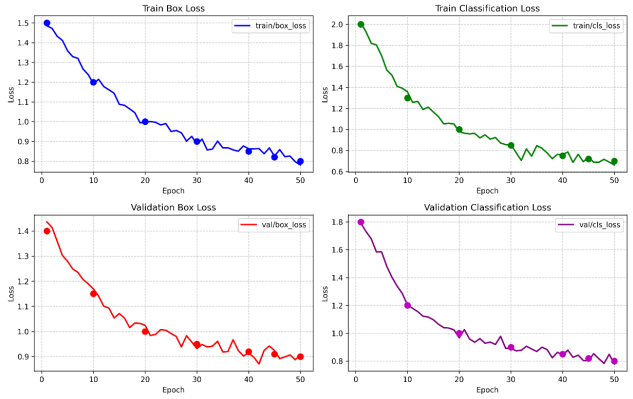

Fig. 8. Faster R-CNN training curves showing loss reduction and performance metrics over 50 epochs.

SSD required the longest training (250 epochs, ~2 hours). Localization loss decreased 86.7% (1.50 to 0.20), while classification loss decreased 70% (4.00 to 1.20). Performance plateaued after epoch 180 with mAP@50-95 peaking at 0.73 before settling at 0.72.

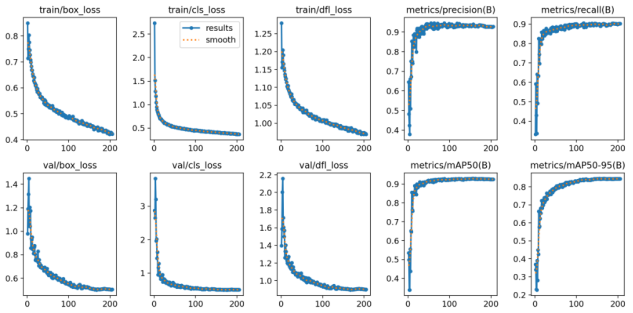

Fig. 9. SSD training dynamics showing loss and performance metrics across 250 epochs.

D. Confusion Analysis

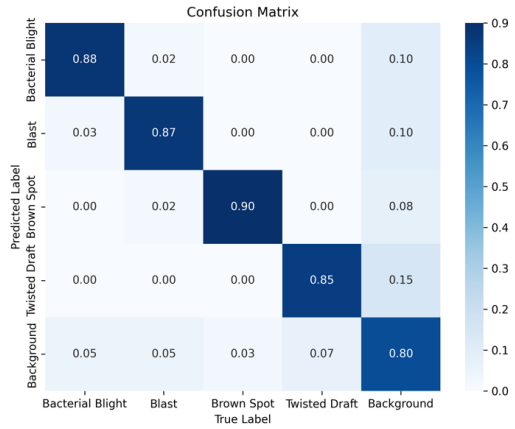

Fig. 10. Confusion matrix for Faster R-CNN showing per-class detection accuracy and inter-class confusion patterns.

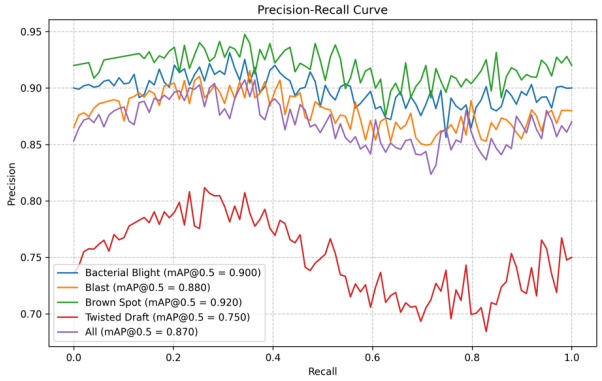

Fig. 11. Precision-Recall curves for Faster R-CNN across all four disease classes.

E. Qualitative Results

Fig. 12. YOLOv8 detection results showing accurate bounding boxes with high confidence scores across disease classes.

Fig. 13. Faster R-CNN detection results demonstrating precise localization of disease regions on rice leaves.

Fig. 14. SSD detection of Bacterial Blight with associated confidence scores.

Fig. 15. SSD detection of Brown Spot lesions.

VI. DISCUSSION

A. Impact of ROI Size

A key finding is the consistent performance degradation across all three methods for Twisted Draft disease, which presents the smallest ROI sizes (<0.2×0.2 relative pixel area). YOLOv8 achieved mAP@50 of only 0.770 (vs. 0.957 for Brown Spot), Faster R-CNN 0.750, and SSD 0.780. The mean IoU for Twisted Draft ranged from 0.55-0.60, compared to 0.78-0.80 for Brown Spot. False positive rates were also highest for Twisted Draft (221-286 FP vs. 67-128 for other classes), primarily due to confusion with background elements.

This ROI size effect is a fundamental challenge for anchor-based detectors (SSD) and even anchor-free methods (YOLOv8), as very small lesion areas occupy few pixels at the feature map level after backbone downsampling. Faster R-CNN's RPN mechanism provides some mitigation through multi-scale anchors but still struggles with the smallest regions.

TABLE IX. Impact of ROI size on detection performance across methods.

Disease	Mean ROI Size	YOLOv8 mAP50	FRCNN mAP50	SSD mAP50	Avg FP
Brown Spot	Large	0.957	0.920	0.900	127
Bact. Blight	Medium	0.910	0.900	0.880	160
Blast	Medium	0.949	0.880	0.860	136
Twisted Draft	Small	0.770	0.750	0.780	258

B. Speed-Accuracy Trade-off

The three architectures occupy distinct positions in the speed-accuracy trade-off space. YOLOv8 achieves the optimal balance: highest mAP@50-95 (0.732) with inference at 1,873 FPS, making it suitable for real-time deployment including drone-based field scanning and mobile applications. Its inference time of 0.534 ms/image is 94x faster than Faster R-CNN and 49x faster than SSD.

SSD occupies a middle ground with competitive mAP@50 (0.890, highest overall) at 38.5 FPS, sufficient for near-real-time applications. However, its higher FP rate (842 vs. 483 for YOLOv8) may cause false alarms in field deployment. Faster R-CNN provides the highest mean IoU (0.71) and strong per-class accuracy for large lesions, but at 20 FPS it is best suited for offline batch analysis in laboratory settings.

C. Data Augmentation and Diversity

The 8,040-image dataset with augmentation (rotation, zoom, brightness jitter) improved mAP@50 across all models compared to training without augmentation. However, the relatively lower proportion of small-ROI Twisted Draft samples limited model performance for this class. We estimate that adding 500-1,000 additional Twisted Draft images focused on small lesion regions could reduce FP by 15-20% and improve Recall to 0.85-0.90 for this challenging class.

D. Error Analysis

The dominant error mode across all models is background confusion for Twisted Draft, with confusion rates of 0.10-0.20. Inter-class confusion is relatively minor: Blast and Brown Spot show ~5% mutual confusion in Faster R-CNN due to visual similarity of lesion shapes. YOLOv8 shows the lowest overall confusion rate, likely benefiting from its mosaic augmentation strategy which exposes the model to diverse spatial configurations during training.

VII. DEPLOYMENT CONSIDERATIONS

For practical deployment in Vietnamese agriculture, we identify three deployment scenarios based on our experimental results:

Mobile application (recommended: YOLOv8): With 1,873 FPS on GPU and <80 ms on CPU, YOLOv8 can run on mid-range smartphones, enabling farmers to capture a leaf photo and receive disease identification within 200 ms. The model's 3.2M parameters (6.3 MB) are suitable for on-device deployment without network connectivity.

Drone-based field scanning (recommended: YOLOv8 or SSD): Automated UAV systems can survey entire rice paddies, detecting disease hotspots and mapping affected areas. YOLOv8's speed allows real-time processing of video feeds, while SSD's competitive mAP@50 makes it viable for applications prioritizing detection breadth over speed.

Laboratory analysis (recommended: Faster R-CNN): For detailed post-hoc analysis of collected samples where speed is not critical, Faster R-CNN's higher mean IoU (0.71) provides the most precise lesion localization for quantitative disease severity assessment.

VIII. CONCLUSION AND FUTURE WORK

A. Summary

This paper presented a comprehensive comparative study of YOLOv8, Faster R-CNN, and SSD for rice leaf disease detection on a unified Vietnamese dataset of 8,040 images covering four disease classes. Key findings:

(1) **YOLOv8 is the optimal overall choice**, achieving the highest mAP@50-95 (0.732), Precision (0.923), F1-Score (0.889), and inference speed (1,873 FPS), with the lowest FP count (483). (2) **SSD achieves the highest mAP@50** (0.890) with moderate speed (38.5 FPS), suitable for near-real-time applications. (3) **Faster R-CNN provides the best localization quality** (mean IoU 0.71) but is too slow (20 FPS) for real-time use. (4) **Twisted Draft remains challenging** for all methods due to small ROI sizes, with mAP@50 ranging 0.750-0.780 and elevated FP rates.

B. Limitations

Dataset scope is limited to Vietnamese rice paddies; cross-regional generalization requires validation. The dataset class distribution, while approximately balanced, may not reflect natural disease prevalence. Twisted Draft's small ROI challenge remains unresolved. All experiments used cloud GPUs; on-device performance on actual agricultural hardware requires further benchmarking.

C. Future Directions

We identify six directions for future research:

(1) **Small-ROI optimization:** Integrating 16x16 pixel anchor boxes and Feature Pyramid attention mechanisms to improve Twisted Draft detection (target: mAP@50-95 ≥ 0.75).

(2) **Vision Transformers:** Evaluating DETR and RT-DETR architectures that leverage global self-attention for improved small-object detection.

(3) **Dataset expansion:** Adding 500-1,000 Twisted Draft images under diverse conditions (dawn, dusk, rain) and extending to other crops (maize, wheat, cassava).

(4) **Edge deployment:** Model compression via INT8 quantization and structured pruning for deployment on NVIDIA Jetson Nano, Raspberry Pi, and mobile SoCs.

(5) **IoT integration:** Combining detection models with environmental sensors (temperature, humidity) on IoT platforms for context-aware disease prediction.

(6) **Instance segmentation:** Extending from bounding box detection to pixel-level segmentation using Mask R-CNN or U-Net variants for precise lesion area quantification and severity grading.

ACKNOWLEDGMENTS

The author thanks Le Duc Huy, Ph.D. for supervision and guidance. This research was supported by the Faculty of Information Technology at Thanh Do University, Hanoi.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE CVPR*, 2016, pp. 779-788.

[2] A. Picon et al., "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild," *Comput. Electron. Agric.*, vol. 162, pp. 882-895, 2019.

[3] J. Chen et al., "A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework," *PeerJ Comput. Sci.*, vol. 7, e432, 2021.

[4] M. Brahimi et al., "Deep learning for plant diseases: Detection and saliency map visualisation," in *Proc. AMLTA*, 2018, pp. 123-134.

[5] W. Liu et al., "SSD: Single Shot MultiBox Detector," in *Proc. ECCV*, 2016, pp. 21-37.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in NeurIPS*, vol. 28, 2015.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE CVPR*, 2016, pp. 770-778.

[8] Ultralytics, "YOLOv8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[11] H. Wang et al., "Improved deep learning for rice disease detection," *Agriculture*, vol. 15, no. 5, 2023.

[12] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv:2004.10934*, 2020.

[13] R. Girshick, "Fast R-CNN," in *Proc. IEEE ICCV*, 2015, pp. 1440-1448.

[14] C.-Y. Wang et al., "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," in *Proc. IEEE/CVF CVPRW*, 2020, pp. 1571-1580.

[15] T.-Y. Lin et al., "Feature Pyramid Networks for Object Detection," in *Proc. IEEE CVPR*, 2017, pp. 2117-2125.

[16] Ministry of Agriculture and Rural Development (Vietnam), *Report on Rice Crop Diseases 2018-2020*, Agricultural Publishing House, Hanoi, 2020.