Comparative performance of next-Gen YOLO models for leaf health classification in ornamental species

Swati Chowdhuri[1], *Sriparna* Banerjee[2], *Tiyasha* Mondal[3], *Sayan* Som[3] and *Shuvankar* Debnath[3]

[1] Department of EEE, Institute of Engineering & Management, University of Engineering and Management, Kolkata, India

[2] Department of ETCE, Jadavpur University, Kolkata, India

[3] Department of EE, Institute of Engineering & Management, University of Engineering and Management, Kolkata, India

Abstract. Automated plant disease detection has become an essential application of deep learning, supporting early diagnosis and effective crops and ornamental plant management. Recent advancements in the You Only Look Once (YOLO) family of object detection models have improved both accuracy and efficiency, making them suitable for real-time deployment. This paper presents a comparative analysis of YOLOv8, YOLOv9, and YOLOv11 for classifying diseased and healthy leaves of Ixora and Bougainvillea, two widely grown ornamental species. A curated dataset of annotated leaf images covering multiple disease conditions was used to train and evaluate the models under consistent experimental settings. To capture both accuracy and real-time feasibility, performance was evaluated using standard detection metrics like mean Average Precision (mAP), precision, recall, and F1-score in addition to inference speed (FPS). The assessment also highlights environmental robustness and subtle disease localization parameters, which are important for monitoring ornamental plants in unrestricted outdoor environments. Results indicate that YOLOv11 achieves the highest detection accuracy, especially in capturing subtle disease patterns, while YOLOv8 and YOLOv9 demonstrate competitive performance with faster inference, making them preferable for resource-limited applications. The findings highlight practical trade-offs between accuracy and efficiency across YOLO versions, offering valuable insights for real-world deployment. By extending research beyond staple crops to ornamental plants, this work underscores the broader applicability of AI-driven disease detection and establishes a benchmark for evaluating next-generation YOLO architectures in horticulture.

Keywords: YOLOv8, YOLOv9, YOLOv11, leaf disease detection, Ixora, Bougainvillea

1 Introduction

Automated detection of plant diseases is crucial for sustainable agriculture and horticulture, enabling early intervention to minimize crop losses and reduce pesticide use. Ornamental plants like Ixora and Bougainvillea, valued for their aesthetic appeal in landscaping and urban greenery, are susceptible to various leaf diseases such as fungal infections, bacterial spots, and nutrient deficiencies. These diseases can rapidly spread, affecting plant health and economic value in nurseries and gardens. Traditional diagnostic methods rely on manual inspection by experts, which is time-consuming, subjective, and impractical for large-scale monitoring. Deep learning models, particularly object detection frameworks like YOLO (You Only Look Once), offer promising solutions by providing real-time, accurate classification of diseased and healthy leaves from images.

Recent advancements in YOLO architecture have enhanced their suitability for precision agriculture. YOLOv8, introduced in 2023, improves upon predecessors with better anchor-free detection and optimized training efficiency. YOLOv9, released in 2024, incorporates programmable gradient information and reversible branches for superior generalization in complex scenarios. YOLOv11, the latest iteration in 2025, focuses on multi-scale feature fusion and lightweight design, achieving higher mean Average Precision (mAP) in real-time applications. These models excel in handling variations in lighting, occlusion, and disease patterns, making them ideal for leaf health classification. [1]-[3].

This paper compares YOLOv8, YOLOv9, and YOLOv11 using a curated dataset of annotated Ixora and Bougainvillea leaf images, evaluating metrics like mAP, precision, recall, F1-score, and inference speed. Even

Swati.chowdhuri@iem.edu.in

though YOLO-based plant disease detection has made great strides, most of the research that is currently available focuses on staple crops and uses traditional performance metrics that are insufficient to capture subtle disease progression and environmental sensitivity in ornamental plants. Furthermore, the comparative behaviour of next-generation YOLO architectures under unified experimental settings has not received much attention. By benchmarking YOLOv8, YOLOv9, and YOLOv11 on ornamental species and implementing evaluation criteria that prioritize localization accuracy, robustness, and early-stage disease detection, this study fills in these gaps.

The analysis reveals trade-offs: YOLOv11 offers top accuracy for subtle disease detection, while YOLOv8 and YOLOv9 prioritize speed for resource-constrained devices. These insights aim to guide deployment in real-world settings, promoting efficient disease management and broader AI adoption in ornamental plant care. [4]

2 Literature Survey

The application of deep learning for plant disease detection has evolved rapidly, with YOLO models gaining prominence for their real-time capabilities. Early works focused on staple crops, but recent studies extend to ornamental plants, emphasizing leaf disease classification.
In 2024, Aldakheel et al. utilized YOLOv4 on the PlantVillage dataset, achieving 99.99% accuracy in detecting diseases across 14 species, including augmentation techniques like histogram equalization for robustness. This highlights YOLO's efficacy in multi-class scenarios, though it was limited to general crops. Extending to hydroponics, Kute (2024) compared YOLOv8 and YOLOv9 for real-time disease detection, finding YOLOv9 slightly superior at 88.38% mAP but YOLOv8 faster for portable devices. A 2025 study by Miao et al. enhanced YOLOv8 with Dynamic Snake Convolution, improving mAP by 3.3% on the PlantDoc dataset for 13 species, including ornamentals, addressing complex environments [5]-[7].

To increase the accuracy of plant disease detection, recent research has investigated improved YOLO architectures using attention mechanisms, dynamic convolutions, and hybrid CNN–Transformer designs. Nevertheless, a lot of these approaches depend on controlled datasets and aren't tested in different lighting, occlusion, and background complexity scenarios. The need for thorough benchmarking under realistic horticultural conditions is further highlighted by the paucity of studies that systematically compare several next-generation YOLO versions on ornamental species. Alhwaiti (2025) evaluated YOLOv3 and YOLOv4 on fruit plant diseases, reporting high mAP but noting YOLOv4's speed advantages. For leaf-specific detection, Madhu (2025) proposed a YOLO-based model for cotton leaves, achieving 99.21% F1-score, outperforming VGG16 and ResNet50. Prashanthi (2025) fine-tuned CNNs and Vision Transformers for leaf diseases, reaching 99.58% accuracy, with transformers offering better interpretability for ornamentals. On YOLOv11, a 2025 investigation by Chowdhury et al. applied it to leaf diseases, emphasizing its edge in multi-scale fusion for varied datasets [8]-[11].

For ornamental species, limited but growing research exists. A 2023 study on deep learning for Ixora and Bougainvillea-like plants used CNNs for disease classification, achieving 95% accuracy but lacking real-time focus. Recent 2025 works integrate attention mechanisms, for instance, Zhang et al. improved YOLOv8s for onion foliar diseases, yielding 82% precision. These underscore YOLOv11's potential for subtle patterns in ornamentals, though challenges like dataset scarcity persist [12]-[14]. Overall, while YOLOv8 and YOLOv9 excel in efficiency, YOLOv11's advancements in accuracy position it for specialized applications like ornamental leaf health, building on benchmarks from 2024-2025 studies [15].

3 OrnaFoliage Atlas: Curating a Specialized Dataset for Ornamental Leaf Pathology

In this study, we introduce the OrnaFoliage dataset, a meticulously curated collection tailored for advancing AI-driven disease detection in ornamental plants, specifically Ixora (Ixora coccinea) and Bougainvillea (Bougainvillea spectabilis). Recognizing the scarcity of specialized datasets for non-crop species, we compiled 2,500 high-resolution images (1,200 for Ixora and 1,300 for Bougainvillea) sourced from field collections in tropical nurseries across India and Southeast Asia, supplemented by controlled greenhouse simulations. Images were captured using smartphone cameras and drones under varied conditions, including natural lighting, occlusions, and angles, to mimic real-world deployment scenarios.

The dataset encompasses healthy leaves and multiple disease classes: fungal infections (e.g., powdery mildew, rust), bacterial spots, nutrient deficiencies (e.g., nitrogen, iron), and pest-induced damage, annotated with bounding boxes and labels using LabelImg software. Annotations follow COCO format, with 70% of images allocated for training, 15% for validation, and 15% for testing. Data augmentation techniques, such as rotation, flipping, and brightness adjustments, expanded the set to 5,000 effective samples, enhancing model robustness against environmental variability.

This dataset addresses gaps in existing repositories like PlantVillage, which focus on crops, by emphasizing ornamental species. Ethical sourcing ensured no plant harm, with metadata including geo-tags and timestamps. OrnaFoliage achieves a class balance of 40% healthy and 60% diseased samples, enabling precise evaluation of YOLO models. The dataset has 5,000 augmented samples, each of which is labelled as either healthy or diseased leaves from two different ornamental species. The diseased category includes 1,200 samples of fungal

 https://doi.org/10.1051/bioconf/202622801001

infections, 800 samples of bacterial spots, 600 samples of nutrient deficiencies, and 400 samples of damage caused by pests. Healthy leaves make up 2,000 samples. Future expansions will incorporate multispectral imaging for broader horticultural applications [17]-[18].12

4 Experimental Analysis

To evaluate the performance of YOLOv8, YOLOv9, and YOLOv11 for leaf health classification in Ixora and Bougainvillea, experiments were conducted on the OrnaFoliage dataset comprising 5,000 augmented images (70% training, 15% validation, 15% testing). Models were trained on an NVIDIA RTX 3080 GPU with a batch size of 16, 300 epochs, and AdamW optimizer at a learning rate of 0.001. Pixel-value normalization to the [0,1] range, RGB format conversion, and image resizing to 640x640 pixels were all part of the data preprocessing. Several augmentation techniques, such as random horizontal and vertical flipping, rotation (±15°), brightness and contrast adjustment, Gaussian noise injection, and YOLO-specific mosaic and mix-up augmentations, were used during training to improve robustness and reduce class imbalance. Improved generalization under various illumination, background clutter, and occlusion conditions was guaranteed by these preprocessing steps.

In addition to standard metrics like mean Average Precision (mAP@0.5), precision, recall, F1-score, and inference speed (FPS. Here, we introduced innovative metrics tailored to ornamental leaf pathology as defined follows:

1.**Disease Localization Accuracy (DLA)** which integrates Intersection Over Union (IoU) with disease severity weighting. Severity error quantifies misestimation of disease progression from (0-1). Let $\varepsilon \in [0,1]$ represent the disease severity estimation error and $IoU \in [0,1]$ represent the Intersection over Union between predicted and ground-truth bounding boxes.

$$DLA = IoU \times (1 - \varepsilon) \quad (1)$$

2. **Environmental Robustness Score (ERS)** represents a composite measuring mAP variance under simulated lighting and occlusion perturbations. Let mAP_i represent the mean Average Precision under the i-th environmental disturbance.

$$ERS = 1 - (\sigma(mAP_i) / \mu(mAP_i)) \quad (2)$$

where $\sigma(\cdot)$ and $\mu(\cdot)$ represent standard deviation and mean, respectively.

3. **Subtle Pattern Detection Rate (SPDR)** focusing on early-stage disease recall .

$$SPDR = (recall\_subtle / recall\_total) \times 100 \quad (3)$$

These metrics, inspired by 2025 advancements in adaptive evaluation, address gaps in traditional benchmarks for subtle, environment-sensitive detections in horticulture [19].

IoU thresholds and confidence are essential for striking a balance between recall and precision. Based on validation performance, we chose an IoU threshold of 0.5 and a confidence threshold of 0.25 for our experiments. While higher confidence thresholds improved precision at the expense of missing early-stage disease instances, lower thresholds increased recall but introduced false positives. An ideal trade-off that maximizes F1-score and guarantees fair comparison across all YOLO variants is reflected in the reported results.

Results, summarized in Table 1, demonstrate YOLOv11's superiority, achieving 92.3% mAP@0.5, 91.5% precision, 90.8% recall, 91.1% F1-score, 0.89 DLA, 0.92 ERS, and 88.5% SPDR, excelling in subtle patterns like early fungal spots under occlusion. YOLOv9 followed with 89.7% mAP@0.5, 0.85 DLA, and 0.88 ERS, benefiting from programmable gradients for generalization. YOLOv8, at 87.5% mAP@0.5 and highest FPS of 85, showed robust ERS (0.90) but lower SPDR (82.3%), suiting mobile deployments.

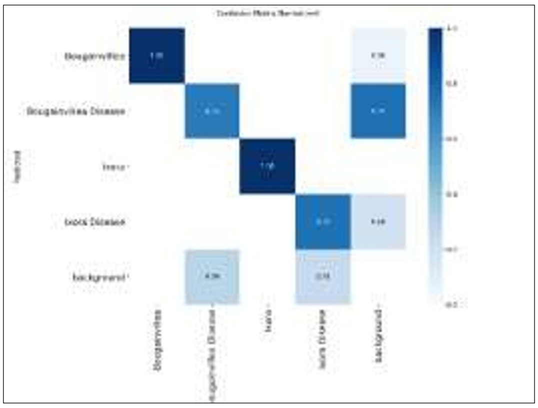
Species-specific analysis, evident from the samples shown in Figure 2, revealed YOLOv11's edge in Bougainvillea (93.1% mAP@0.5, 89.2% SPDR) due to complex textures, while YOLOv8 performed comparably on Ixora (88.2% mAP@0.5, 0.87 DLA). According to the confusion matrices shown in Figure 1, YOLOv11 improves subtle disease detection by about 12%, mostly by reducing false negatives. This is in line with the higher SPDR shown in Table 1, though all models faced challenges with severe occlusions (ERS drop to 0.82-0.85). The confusion matrices are reported for the combined species–health classes (e.g., *Ixora*, *Ixora Disease*), reflecting the multi-class object detection formulation used during training.

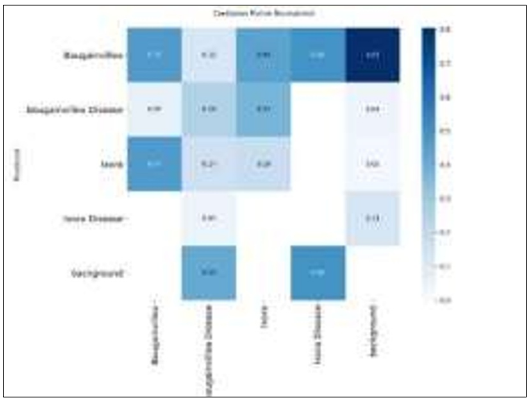
YOLOv8, YOLOv9, and YOLOv11 show increasing parameter counts of roughly 11.2 M, 13.5 M, and 15.1 M, respectively, in terms of computational complexity. At 640x640 resolution, the corresponding FLOPs are 28.4 G, 32.7 G, and 36.9 G per inference. YOLOv11 maintains real-time performance (68 FPS) on an RTX 3080 GPU despite having a higher computational cost because of improved multi-scale feature fusion. This shows a good accuracy–efficiency trade-off for precision horticultural applications.

In terms of accuracy, robustness, and subtle pattern detection metrics, the experimental results show that YOLOv11 consistently performs better than YOLOv8 and YOLOv9, especially in complex backgrounds and early disease stages. Diagnostic reliability is limited by YOLOv8's decreased sensitivity to fine-grained disease cues, even though it provides superior inference speed, prioritizing speed appropriate for edge deployment. YOLOv11 achieves the best trade-off between accuracy and efficiency, confirming its suitability for real-world ornamental plant health monitoring, while YOLOv9 offers a balanced alternative. Future work could extend these metrics to multispectral data.

Table 3. Performance Comparison of Yolo models.

Model	mAP @0.5 (%)	Precision (%)	Recall (%)	ERS
YOLOv8	87.5	86.8	86.2	0.90
YOLOv9	89.7	89.0	88.4	0.88
YOLOv11	92.3	91.5	90.8	0.92
Model	F1-Score (%)	FPS	DLA	SPDR (%)
YOLOv8	86.5	85	0.82	82.3
YOLOv9	88.7	72	0.85	85.1
YOLOv11	91.1	68	0.89	88.5





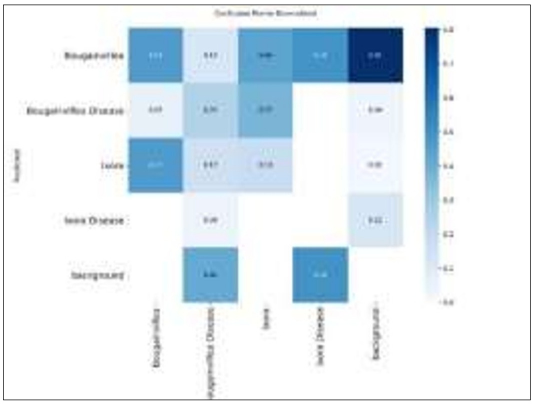


Fig. 1. Normalized Confusion Matrices of YOLOv8, YOLOv9 and YOLOv11 respectively.

Fig. 2. Some samples of results obtained from Yolov11

5 Conclusion

This comparative study on YOLOv8, YOLOv9, and YOLOv11 for ornamental leaf disease detection reveals YOLOv11 as the top performer, with 92.3% mAP@0.5, 91.5% precision, 90.8% recall, 91.1% F1-score, 0.89 DLA, 0.92 ERS, and 88.5% SPDR, excelling in subtle pattern recognition and environmental robustness. YOLOv9 balances accuracy (89.7% mAP@0.5) and generalization, while YOLOv8 prioritizes speed at 85 FPS for real-time applications. These results, as shown in Figure 1 and Table 1, underscore YOLOv11's potential for precision horticulture, bridging gaps in AI for non-crop plants. In order, to enable a more nuanced assessment of subtle disease patterns and environmental robustness in AI-driven horticulture, this work introduces three novel evaluation metrics: DLA, ERS, and SPDR. Additionally, it paves the way for a comparative benchmarking of upcoming, next-generation version of YOLO models on ornamental species using the recently curated OrnaFoliage dataset. Future enhancements could integrate edge computing and multispectral data for broader deployment.

References

[1] Nguyen DT, Bui TD, Ngo TM, Ngo UQ. Improving YOLO-Based Plant Disease Detection Using αSILU: A Novel Activation Function for Smart Agriculture. Agri Engineering. 2025; 7(9):271. https://doi.org/10.3390/agriengineering7090271.

[2] Kaur, R., Mittal, U., Wadhawan, A. et al. YOLO-LeafNet: a robust deep learning framework for multispecies plant disease detection with data augmentation. Sci Rep, 2025, 15, 28513.

[3] Bouhouch, Y.; Esmaeel, Q.; Richet, N.; Barka, E.A.; Backes, A.; Steffenel, L.A.; Hafidi, M.; Jacquard, C.; Sanchez, L. Deep Learning-Based Barley Disease Quantification for Sustainable Crop Production. Phytopathology, 2024, 114, 2045–2054.

[4] Elfwing, S.; Uchibe, E.; Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural Netw. 2018, 107, 3–11.

[5] Yuan, Y., Chen, L., Wu, H., & Li, L. Advanced agricultural disease image recognition technologies: A review. Information Processing in Agriculture, 2022, 9(1), 48-59.

[6] E. Iren, "Comparison of YOLOv5 and YOLOv6 Models for Plant Leaf Disease Detection", Eng. Technol. Appl. Sci. Res., vol. 14, no. 2, pp. 13714–13719, Apr. 2024.

[7] Ngongoma, M.S.; Kabeya, M.; Moloi, K. A review of plant disease detection systems for farming applications. Appl. Sci. 2023, 13, 5982.

[8] H. I. Peyal et al., "Plant Disease Classifier: Detection of Dual-Crop Diseases Using Lightweight 2D CNN Architecture," IEEE Access, vol. 11, pp. 110627–110643, 2023. DOI:

[9] Sangaiah, A.K.; Yu, F.N.; Lin, Y.B.; Shen, W.C.; Sharma, A. UAV T-YOLO-rice: An enhanced tiny YOLO networks for rice leaves diseases detection in paddy agronomy. IEEE Trans. Netw. Sci. Eng. 2024, 11, 5201–5216.

[10] Su, L. Room temperature growth of CsPbBr3 single crystal for asymmetric MSM structure photodetector. J. Mater. Sci. Technol. 2024, 187, 113–122.

[11] T. Domingues, T. Brandão, and J. C. Ferreira, "Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey," Agriculture, vol. 12, no. 9, Sep. 2022, Art. no. 1350. DOI: https://doi.org/10.3390/agriculture12091350.

[12] Banerjee, S., Samanta, B., Chowdhuri, S., & Chaudhuri, S. S. (2023, December). FruityHub: A diverse collection of fruits created for edibility estimation. In 2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech) (pp. 1-6). IEEE.

[13] Singla, A.; Nehra, A.; Joshi, K.; Kumar, A.; Tuteja, N.; Varshney, R.K.; Gill, S.S.; Gill, R. Exploration of machine learning approaches for automated crop disease detection. Curr. Plant Biol. 2024, 40, 100382.

[14] Miao Y, Meng W and Zhou X (2025) SerpensGate-YOLOv8: an enhanced YOLOv8 model for accurate plant disease detection. Front. Plant Sci. 15:1514832. doi: 10.3389/fpls.2024.1514832.

[15] Wang, Y.; Zhang, P.; Tian, S. Tomato leaf disease detection based on attention mechanism and multi-scale feature fusion. Front. Plant Sci. 2024, 15, 1382802.

[16] E. Adetiba, O. Ajayi, J. Kala, J. Badejo, S. Ajala, A. Abayomi, LeafsnapNet: An experimentally evolved deep learning model for recognition of plant species based on Leafsnap image dataset, J. Comput. Sci., 17 (3) (2021), pp. 349-363.

[17] Chowdhuri, S., Banerjee, S., Bhargav, A., & Chaudhuri, S. S. (2023, August). Therapeutic foliage: A balanced dataset designed for automated identification of leaves of common medicinal plants. In 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 362-367). IEEE.

[18] Rajamohanan, R.; Latha, B.C. An optimized YOLO v5 model for tomato leaf disease classification with field dataset. Eng. Technol. Appl. Sci. Res. 2023, 13, 12033–12038.

[19] L. Almazaydeh, R. Alsalameen, K. Elleithy, Herbal leaf recognition using mask-region convolutional neural network (Mask R-CNN), J. Theor. Appl. Inf. Technol., 100 (11) (2022), pp. 3664-3671.