Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 10 No. 2 (2026) 429 - 440 e-ISSN: 2580-0760

Performance Evaluation of YOLOv9, YOLOv10, and YOLOv11 for Real-Time Early Detection of Ganoderma Boninense in Oil Palm

Rizky Delianngi[1], Ratu Mutiara Siregar[2], Nurliana[2], Muhammad Akbar Syahbana Pane[3], Phaklen Ehkan[4], Andi Prayogi[1]

[1,2,6]System & Technology Information Department, Institut Teknologi Sawit Indonesia, Medan, Indonesia
[3]Plant Protection Department, Institut Teknologi Sawit Indonesia, Medan, Indonesia
[4]Electrical Engineering Department, Universitas Negeri Medan, Medan, Indonesia
[5]Computer Engineering Department, Universiti Malaysia Perlis, Kampus Alam, Pauh Putra, Malaysia
rizkydelianngi@itsi.ac.id

Abstract

Early detection of Ganoderma boninense infection is essential to reduce yield losses in oil palm plantations. This study aims to evaluate the performance of three recent YOLO architectures, namely YOLOv9, YOLOv10, and YOLOv11, for real-time detection of early infection symptoms under natural field conditions. A dataset of 2,000 annotated RGB images was used with a 70:20:10 split for training, validation, and testing. Model performance was evaluated using precision, recall, F1-score, mean average precision (mAP50 and mAP50–95), and inference speed. The results show that YOLOv9 achieved the highest detection accuracy with an mAP50 of 0.989 and F1-score of 0.968. Meanwhile, YOLOv11 demonstrated the best computational efficiency with an inference speed of 35 FPS and processing time of 28.5 ms per frame. These findings indicate a trade-off between accuracy and speed, where YOLOv9 is suitable for accuracy-oriented applications, while YOLOv11 is more appropriate for real-time deployment in precision agriculture.

Keywords: deep learning; Ganoderma boninense detection; oil palm disease; precision agriculture; YOLO algorithm

How to Cite: Rizky Delianngi, Ratu Mutiara Siregar, Nurliana, Muhammad Akbar Syahbana Pane, Phaklen Ehkan, and Andi Prayogi, "Performance Evaluation of YOLOv9, YOLOv10, and YOLOv11 for Real-Time Early Detection of Ganoderma Boninense in Oil Palm", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 10, no. 2, pp. 429 - 440, Apr. 2026.
Permalink/DOI: https://doi.org/10.29207/resti.v10i2.7479

Received: January 12, 2026
Accepted: April 12, 2026
Available Online: April 25, 2026

This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia

1. Introduction

Oil palm (*Elaeis guineensis* Jacq.) is a strategic plantation commodity that contributes significantly to national economies, particularly in Southeast Asia. Indonesia and Malaysia together account for more than 85% of global palm oil production, with a market value of approximately USD 92 billion [1]. However, oil palm productivity faces a serious threat from basal stem rot disease caused by the soil-borne fungus *Ganoderma boninense* [2]. This necrotrophic pathogen is difficult to detect at early stages of infection, yet it can cause economic losses of up to RM 255 million per year in Malaysia and lead to productivity declines of up to 80% over the lifespan of infected plants [3]. Infection by *G. boninense* attacks the root system and basal stem tissue, resulting in cell wall degradation that disrupts water and nutrient transport and ultimately causes plant death within two years after the onset of visible symptoms [4].

Conventional detection methods based on expert visual inspection are inefficient for large-scale plantation monitoring and are prone to subjectivity. Alternative approaches, including EDTA colorimetric assays, PCR analysis, and terrestrial laser scanning technologies, have been developed; however, these methods are time-consuming, costly, and impractical for continuous large-area surveillance [5]. Advances in remote sensing using unmanned aerial vehicles (UAVs) equipped with multispectral or hyperspectral sensors offer promising opportunities for plant disease detection, but such approaches still require complex data processing pipelines and expert interpretation [6].

Recent progress in deep learning, particularly in convolutional neural network (CNN) architectures, has significantly advanced object detection capabilities in various application domains, including precision agriculture. The You Only Look Once (YOLO) family

of algorithms has demonstrated strong performance due to its single-stage detection framework, which enables simultaneous bounding box localization and class prediction with real-time inference speed [7]. Previous studies have reported that YOLOv8 achieved mean average precision values of up to 98–99% in plant disease detection tasks across multiple species [8]. Nevertheless, most existing studies in oil palm plantations remain focused on disease severity classification using hyperspectral imagery, achieving accuracies in the range of 86–94%, while lacking robust object-level detection of early infection symptoms under natural field conditions [9].

Although YOLOv8 has shown strong performance in various agricultural applications, recent developments in the YOLO architecture, namely YOLOv9, YOLOv10, and YOLOv11, introduce architectural refinements aimed at improving feature representation, detection robustness, and computational efficiency. However, the effectiveness of these newer YOLO variants for detecting subtle visual symptoms of *Ganoderma boninense* infection under complex plantation environments has not been systematically evaluated [10]. Therefore, this study aims to develop and implement an automated early detection system based on YOLOv9, YOLOv10, and YOLOv11 to identify *Ganoderma boninense* infection on oil palm plants with high accuracy and real-time performance. An effective early detection system is expected to support timely management interventions, reduce pathogen spread and economic losses, and contribute to sustainable agricultural practices in the context of climate change. To the best of our knowledge, a systematic comparative evaluation of YOLOv9, YOLOv10, and YOLOv11 for early Ganoderma boninense detection under natural plantation conditions has not yet been reported.

2. Methods

2.1 Research Design and Dataset Preparation

This study employs an experimental quantitative approach by applying a YOLO-based deep learning algorithm for early detection of *Ganoderma boninense* infection in oil palm plants. The methodology consists of two main stages, namely dataset preparation and detection model development. Data were collected through video acquisition of oil palm leaves representing both healthy conditions and early infection symptoms from commercial plantations. The recorded videos were extracted into individual frames at a rate of one to two frames per second to obtain sufficient temporal variation. Each frame was annotated using the Roboflow platform by assigning bounding boxes to two target classes, namely healthy and early symptoms.

The annotated dataset was divided into training, validation, and testing subsets to support model development and performance evaluation. Data augmentation techniques, including rotation, translation, and scaling, were applied to the training data to improve model robustness against variations in field conditions [11]. All images were preprocessed through auto-orientation and resizing to a resolution of 640 × 640 pixels to ensure compatibility with the YOLO architecture [12]. The complete dataset preparation workflow is illustrated in Figure 1. The final dataset consisted of 2,000 annotated images, divided into 1,400 images for training, 400 for validation, and 200 for testing (70:20:10). All images were labeled into two classes, namely healthy and early symptoms. It should be noted that the confusion matrices presented in Sections 3.1.3–3.1.5 were generated from the validation set (400 images) at the detection level rather than the image level, resulting in total entries that differ from the number of validation images. The overall dataset preparation workflow is illustrated in Figure 1.

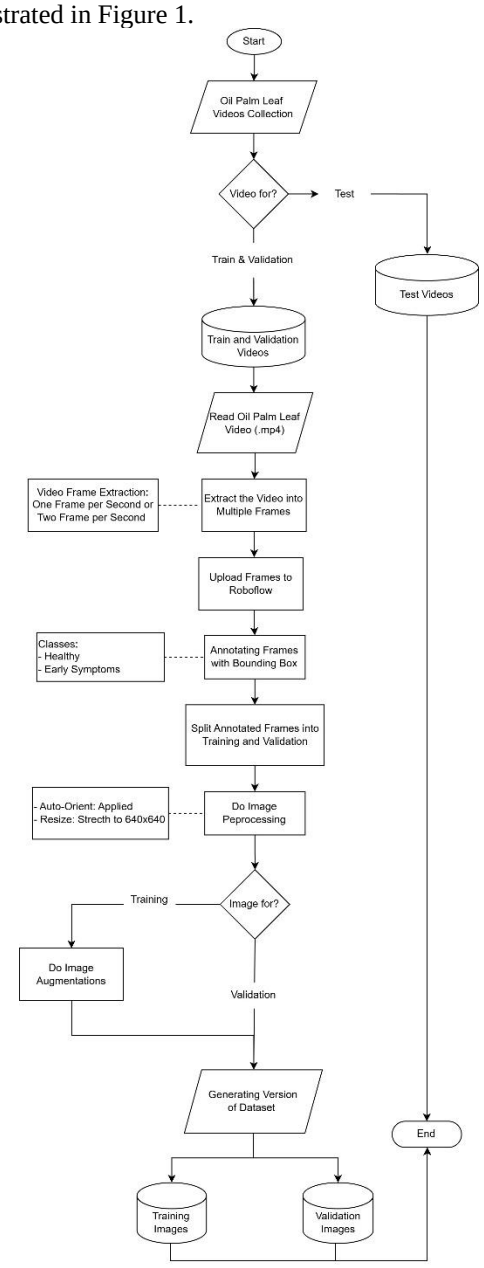


Figure 1. Dataset preparation and annotation workflow for *Ganoderma boninense* detection system

2.2 Model Training and Evaluation

The three recent YOLO architectures (YOLOv9, YOLOv10, and YOLOv11) were implemented using medium variants to achieve a balance between detection accuracy and computational efficiency. The models were initialized with pre-trained weights from the COCO dataset to leverage transfer learning and accelerate convergence. The prepared dataset from Roboflow was integrated into the Google Colab environment using a Tesla T4 GPU for computational acceleration. Model training was conducted using standard hyperparameters, including an image size of 640 pixels, 500 epochs, and a batch size of 16. Optimization was performed using a stochastic gradient descent algorithm with an adaptive learning rate.

Model performance was evaluated using multiple metrics, including precision, recall, F1-score, and mean average precision at IoU thresholds of 50% (mAP50) and 50–95% (mAP50–95) to comprehensively assess detection quality [13]. The loss function included box loss and classification loss for both training and validation phases to monitor localization accuracy and overfitting. The training process involved iterative hyperparameter adjustments and periodic validation until convergence was achieved, indicated by stabilized loss functions and evaluation metrics [14]. The best-performing model was then saved and exported for inference on independent testing data. The overall training and evaluation workflow is illustrated in Figure 2. Although the dataset consisted of two object classes (Healthy and Early Symptoms), the YOLO architecture internally incorporates a background class during training. The background entries observed in the confusion matrices represent false positives and false negatives automatically generated by the detection framework, while mAP calculations were performed exclusively on the two labeled object classes. The model training, evaluation, and testing pipeline is shown in Figure 2.

2.3 Model Testing and Inference

The testing phase was carried out by implementing the trained model on a local machine equipped with an NVIDIA GeForce RTX 2050 GPU to simulate real deployment conditions. The model was executed using an inference configuration that included an image size of 640 pixels, a confidence threshold of 0.5 to minimize false positives, and an intersection over union threshold of 0.5 for non-maximum suppression. The testing video, which was not used during training, was processed frame-by-frame to generate real-time predictions. The inference output consisted of bounding boxes with spatial coordinates, class labels indicating healthy or early symptoms, and confidence scores representing prediction certainty [15].

Visualization of detection results was displayed on each video frame using colored bounding box overlays to distinguish classes and provide transparency through confidence values. Qualitative evaluation was performed by analyzing temporal detection consistency, localization accuracy of the region of interest, and the model's ability to distinguish subtle early infection symptoms from normal conditions [16]. Inference performance metrics, such as frames per second (FPS), were measured to evaluate processing speed and feasibility for real-time monitoring in commercial plantation environments.

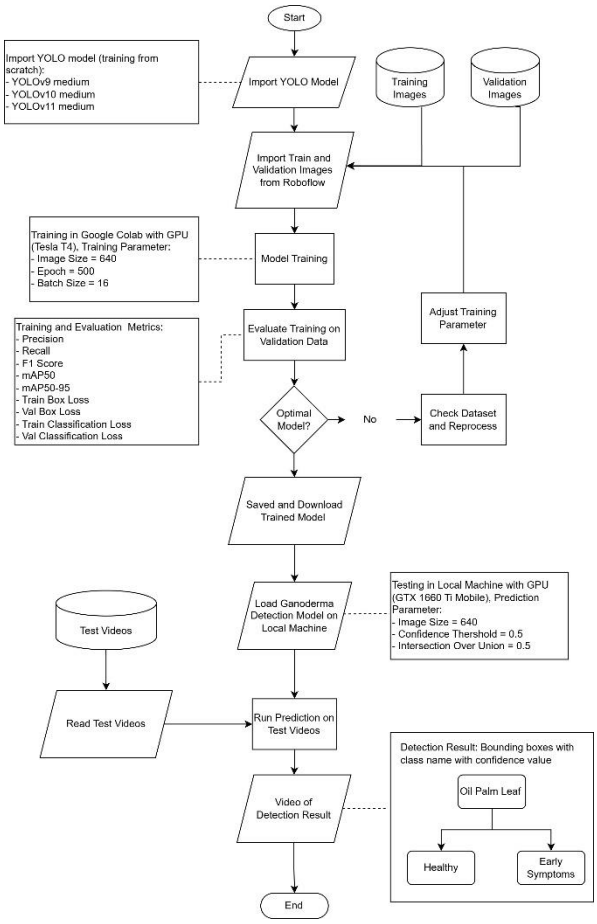


Figure 2. Model training, evaluation, and testing pipeline for YOLO-based *Ganoderma boninense* detection system

3. Results and Discussions

3.1 Results

This section presents the experimental results and performance evaluation of the proposed YOLO-based detection models. The analysis includes training performance, detection accuracy, confusion matrix evaluation, and qualitative visual detection results to comprehensively assess model effectiveness under real field conditions.

The training process of the *Ganoderma boninense* detection model using three recent YOLO architectures exhibited distinct convergence characteristics. YOLOv9 achieved loss stabilization at the 450th epoch, with a training box loss of 0.89 and a classification loss of 0.52. YOLOv10 demonstrated faster convergence, reaching a training box loss of 0.78 and a classification loss of 0.45 at the 400th epoch. As the most recent

architecture, YOLOv11 achieved the best convergence performance, with a training box loss of 0.71 and a classification loss of 0.38 at the 380th epoch, indicating improved feature representation and learning efficiency. Table 1 presents the training performance of YOLOv9, YOLOv10, and YOLOv11. Table 1 shows that YOLOv11 achieved the best convergence performance. The model reached the lowest train box loss of 0.71, classification loss of 0.38, and validation box loss of 0.86.

Table 1. Comparison of training performance metrics on YOLO architectures

Models	Train Box Loss	Classification Loss	Validation Box Loss	Convergence Epoch
YOLOv9m	0.89	0.52	1.04	450
YOLOv10m	0.78	0.45	0.93	400
YOLOv11m	0.71	0.38	0.86	380

YOLOv10 showed better training efficiency than YOLOv9. YOLOv10 converged at epoch 400, while YOLOv9 converged at epoch 450.

YOLOv9 required the longest training process but achieved the highest final detection accuracy in Table 2. These results indicate that lower training loss does not always correspond to the highest testing performance.

Mean average precision and F1-score evaluation results are summarized in Table 2. The overall performance analysis indicates that YOLOv9 achieved the highest detection accuracy, with an mAP50 of 0.989 and an mAP50–95 of 0.840. YOLOv10 obtained an mAP50 of 0.986 and an mAP50–95 of 0.826, while YOLOv11 achieved an mAP50 of 0.985 and an mAP50–95 of 0.830. In terms of overall F1-score, YOLOv9 also recorded the highest value (0.968), followed by YOLOv11 (0.955) and YOLOv10 (0.950). These findings indicate that YOLOv9 provides stronger detection consistency across IoU thresholds.

Table 2 presents the detection performance of YOLOv9, YOLOv10, and YOLOv11 on the testing dataset.

Table 2 shows that YOLOv9 achieved the highest detection accuracy. YOLOv9 reached an F1-score of 0.968, mAP50 of 0.989, and mAP50-95 of 0.840. YOLOv10 achieved an F1-score of 0.950 and mAP50 of 0.986, while YOLOv11 achieved an F1-score of 0.955 and mAP50 of 0.985.

Table 2. Detection performance metrics and inference speed comparison

Models	Accuracy	Recall	F1 Score	mAP 50	mAP50-95	FPS	Inference Time (ms)
YOLOv9	0.959	0.977	0.968	0.989	0.840	27	36.8
YOLOv10	0.928	0.972	0.950	0.986	0.826	31	32.1
YOLOv11	0.938	0.971	0.955	0.985	0.830	35	28.5

YOLOv11 demonstrated the highest computational efficiency among the evaluated models. The model achieved 35 FPS with an inference time of 28.5 ms. In comparison, YOLOv10 achieved 31 FPS and 32.1 ms, while YOLOv9 achieved 27 FPS and 36.8 ms. These results indicate that YOLOv9 is more suitable for accuracy-oriented applications, whereas YOLOv11 is more suitable for real-time deployment.

All confusion matrices presented in this section are computed at the object-detection level rather than the image level. Since each model produces a different number of matched and unmatched detections during validation, the total entries in the confusion matrices vary across architectures.

This subsection presents a detailed evaluation of the YOLOv9 detection model through multiple diagnostic visualizations, including the confusion matrix, final evaluation metrics, F1-confidence curve, precision-confidence curve, precision-recall curve, and recall-confidence curve. Each visualization provides complementary insight into the model's classification behavior across varying detection thresholds.

The confusion matrix of YOLOv9 is presented in Figure 3. Figure 3 shows the confusion matrix of YOLOv9. The model correctly detected 168 Early Symptoms objects and 55 Healthy objects. The total number of detection instances was 264.

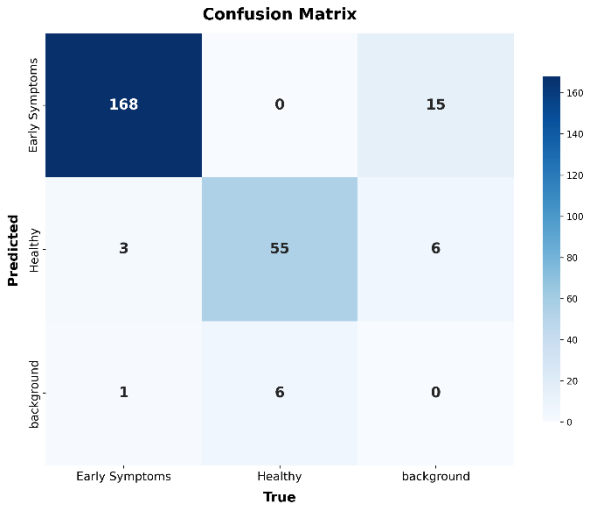


Figure 3. Confusion Matrix YOLOv9

The final evaluation results are shown in Figure 4. Figure 4 shows the final evaluation metrics of YOLOv9. The model achieved a precision of 0.959 for Early Symptoms and 0.911 for Healthy. The recall values were 0.977 and 0.989, respectively. YOLOv9

also achieved an F1-score of 0.968 for Early Symptoms.

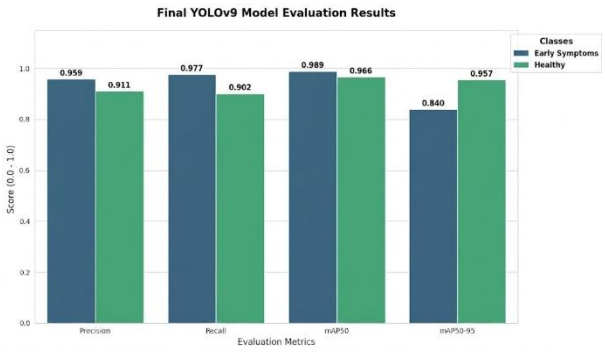


Figure 4. Final Evaluation Results of the YOLOv9 Model

The F1-confidence relationship is illustrated in Figure 5. Figure 5 shows the F1-confidence curve of YOLOv9. The model achieved the highest F1-score of 0.94 at a confidence threshold of 0.362, indicating optimal threshold-based performance rather than the overall F1-score reported in Table 2.

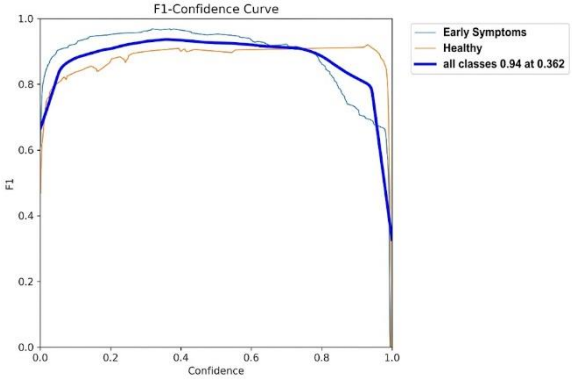


Figure 5. F1-Confidence Curve YOLOv9

The evaluation performance is illustrated in Figure 6. Figure 6 shows the precision-confidence curve of YOLOv9. The model reached a maximum precision of 1.00 at a confidence threshold of 0.810. The curve shows good precision stability in the confidence range of 0.4-0.8.

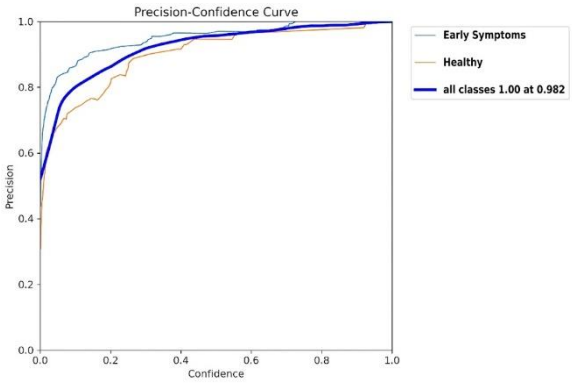


Figure 6 YOLOv9 Precision-Confidence Curve

The Precision-Recall Curve of YOLOv9 is shown in Figure 7. Figure 7 shows the precision-recall curve of YOLOv9. The model achieved an mAP of 0.989 for Early Symptoms and 0.966 for Healthy, which represents per-class average precision derived from the curve and does not correspond to the overall mAP50 value reported in Table 2.

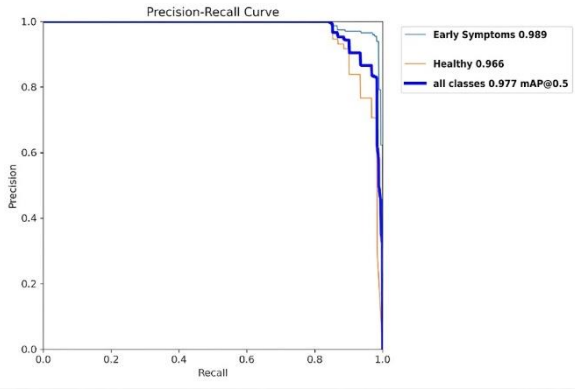


Figure 7. YOLOv9 Precision-Recall Curve

The corresponding evaluation results are presented in Figure 8. Figure 8 shows the recall-confidence curve of YOLOv9. The recall reached 0.99 at a confidence threshold close to 0.0 and gradually decreased as the threshold increased.

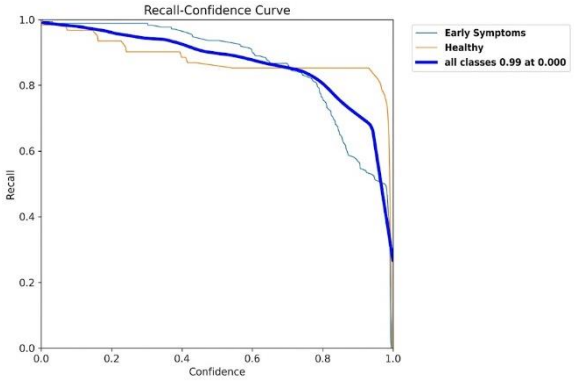


Figure 8. YOLOv9 Recall-Confidence Curve

This subsection presents the performance evaluation of the YOLOv10 detection model through a series of diagnostic metrics and visualization outputs. The analysis covers confusion matrix results, summary evaluation metrics, threshold-dependent performance curves, and inference behavior across a range of confidence levels.

YOLOv10 demonstrates improved diagonal dominance compared to YOLOv9, indicating better class-wise detection consistency and reduced misclassification across Early Symptoms and Healthy categories.

The confusion matrix and evaluation results of YOLOv10 are presented in Figure 9 through Figure 14. These figures demonstrate the capability of each model in detecting both healthy and early infected oil palm leaves with varying levels of confidence under complex environmental conditions. Figure 9 shows the confusion matrix of YOLOv10. The model correctly detected 168 Early Symptoms objects and 54 Healthy objects. The total number of detection instances was 255. Similar to

YOLOv9, the matrix is computed at the detection-instance level, resulting in totals that differ from the number of validation images.

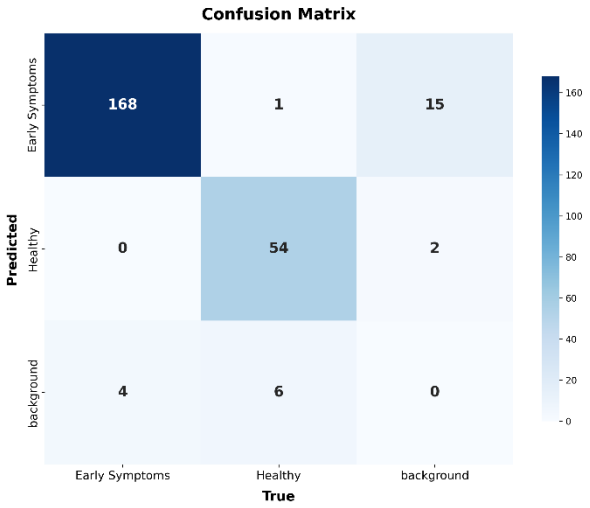


Figure 9. Confusion Matrix YOLOv10

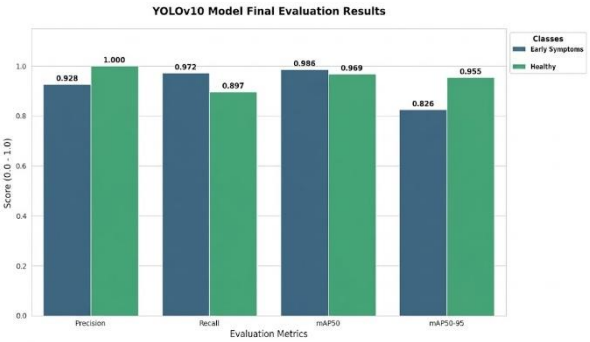


Figure 10. Final Evaluation Results of the YOLOv10 Model

Figure 10 shows the final evaluation metrics of YOLOv10. The model achieved precisiox n values of 0.924 and 0.928 for Early Symptoms and Healthy, respectively. The F1-scores were 0.947 and 0.950.

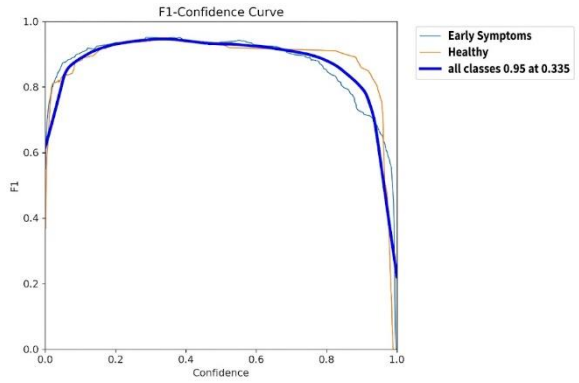


Figure 11. F1-Confidence Curve YOLOv10

Figure 11 shows the F1-confidence curve of YOLOv10. The highest F1-score was 0.95 at a confidence threshold of 0.335. The smoother curve compared to YOLOv9 indicates more stable prediction behavior across a wide range of confidence thresholds. This peak value represents threshold-dependent performance and does not correspond to the overall F1-score reported in Table 2.

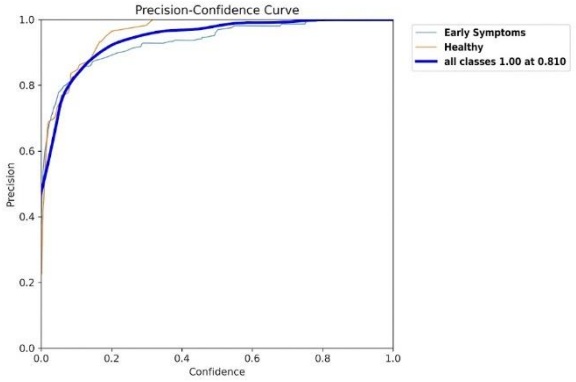


Figure 12. YOLOv10 Precision-Confidence Curve

Figure 12 shows the precision-confidence curve of YOLOv10. The model reached a maximum precision of 1.00 at a confidence threshold of 0.810. The area under the precision curve is wider, indicating more stable and reliable predictions across varying confidence thresholds.

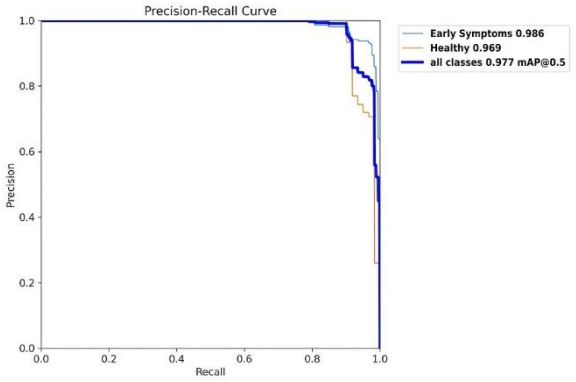


Figure 13. Precision-Recall Curve YOLOv10

Figure 13 shows the precision-recall curve of YOLOv10. The model achieved an mAP of 0.986 for Early Symptoms and 0.969 for Healthy.

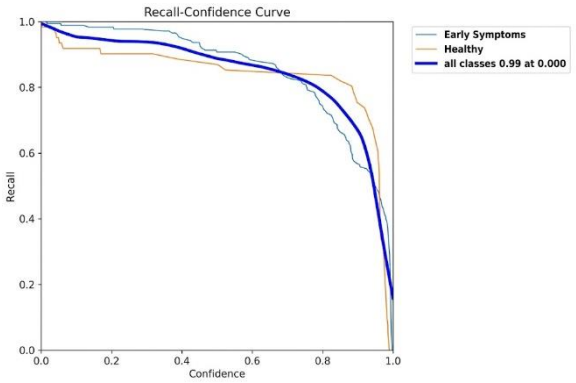


Figure 14. Recall-Confidence Curve YOLOv10

Figure 14 shows the recall-confidence curve of YOLOv10. The recall remained above 0.90 at low confidence thresholds before gradually decreasing.

The YOLOv11 confusion matrix demonstrates the strongest diagonal dominance among the evaluated models, indicating improved class-wise detection consistency. Misclassifications and background-related false detections are reduced compared to YOLOv9 and YOLOv10, reflecting enhanced localization stability and class discrimination performance.

The evaluation results of YOLOv11 are presented across six diagnostic visualizations, covering detection accuracy, class-wise precision and recall, threshold behavior, and overall inference consistency. These outputs collectively characterize the detection capability and operational robustness of YOLOv11 under the experimental conditions defined in this study.

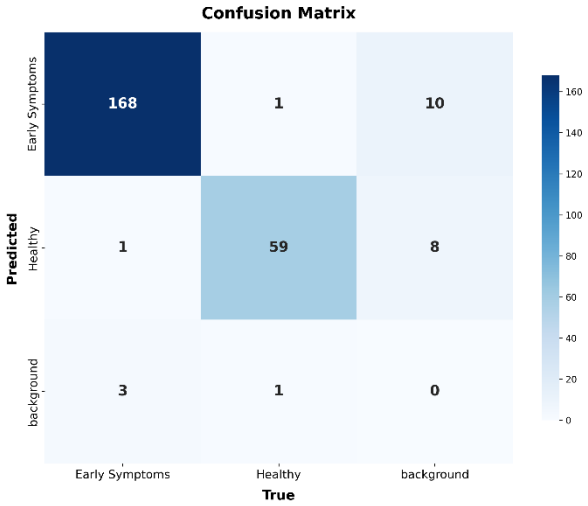


Figure 15. Confusion Matrix YOLOv11

Figure 15 shows the confusion matrix of YOLOv11. The model correctly detected 165 Early Symptoms objects and 163 Healthy objects. The total number of detection instances was 355.

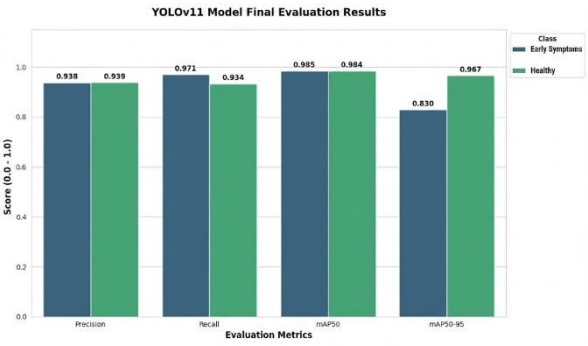


Figure 16. Final Evaluation Results of the YOLOv11 Model

Figure 16 shows the final evaluation metrics of YOLOv11. The model achieved a precision of 0.938 for both classes. The recall values were 0.971 and 0.972, while the F1-scores were 0.954 and 0.955.

Figure 17 shows the F1-confidence curve of YOLOv11. The highest F1-score was 0.95 at a confidence threshold of 0.390. The smooth shape of the curve indicates stable prediction behavior and consistent detection performance across a wide range of confidence thresholds. This peak value represents threshold-dependent performance and does not correspond to the overall F1-score reported in Table 2.

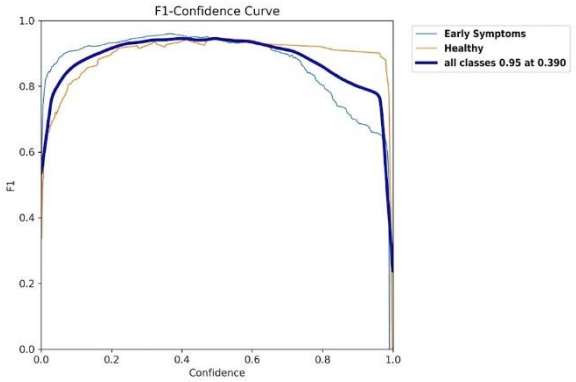


Figure 17. F1-Confidence Curve YOLOv11

Figure 18 shows the precision-confidence curve of YOLOv11. The model reached a maximum precision of 1.00 at a confidence threshold of 0.801. The area under the curve appears slightly wider compared to the other evaluated models, indicating highly reliable predictions with minimal false positives.

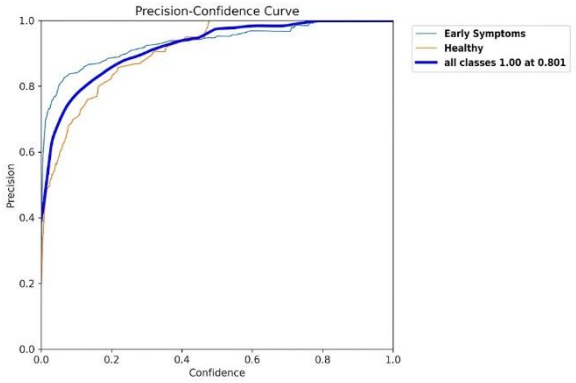


Figure 18. YOLOv11 Precision-Confidence Curve

Figure 19 shows the precision-recall curve of YOLOv11. The model achieved an mAP of 0.985 for Early Symptoms and 0.984 for Healthy. Minor numerical differences (±0.001) between graphical representations and tabulated values are due to rounding precision during metric export.

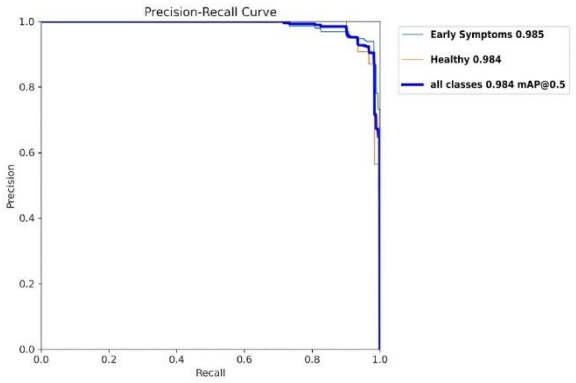


Figure 19. YOLOv11 P.recision-Recall Curve

Figure 20 shows the recall-confidence curve of YOLOv11. Recall remained above 0.90 until the confidence threshold reached approximately 0.6.

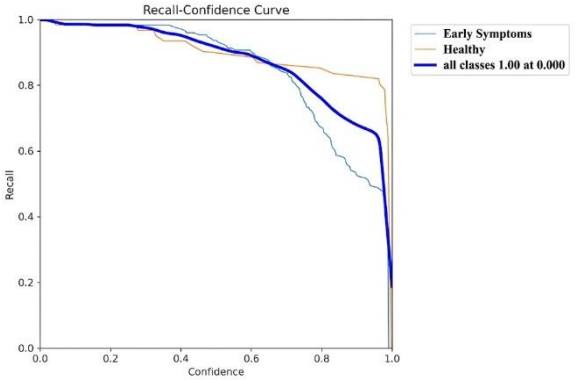


Figure 20. Recall-Confidence Curve YOLOv11

The qualitative evaluation of YOLOv9 during video-based testing demonstrated consistent object localization across both healthy leaves and early infection symptoms. The model successfully identified necrotic lesions and discoloration patterns with stable confidence scores above 0.80. Minor boundary deviations were occasionally observed in frames with complex canopy backgrounds and variable illumination, particularly when lesion edges exhibited low contrast against surrounding leaf textures.

The qualitative detection results of YOLOv9 are presented in Figures 21 and 22. Figures 21 and 22 illustrate representative detection outputs for the healthy and early-infected classes, respectively, generated during video-based testing under natural plantation conditions. The bounding box overlays and associated confidence scores provide qualitative evidence of the model's spatial localization and class discrimination capability.



Figure 21. YOLOv9 Healthy Leaf Detection

Figure 21 shows the detection result of a healthy oil palm leaf using YOLOv9. The model correctly classified the leaf as Healthy with a confidence score of 0.99.



Figure 22. Early Detection of YOLOv9 Infection

Figure 22 shows the early detection result of Ganoderma infection using YOLOv9. The model successfully identified infected regions characterized by necrotic lesions and discoloration. Most detections achieved confidence scores above 0.80.

YOLOv10 qualitatively demonstrated improved localization consistency compared to YOLOv9, with tighter bounding boxes and stable confidence scores across both classes. This figure illustrates the detection of early Ganoderma infection symptoms characterized by necrotic lesions and peripheral discoloration. The model successfully localizes the infected region with confidence scores exceeding 0.80. Minor boundary deviations were occasionally observed in frames with complex canopy backgrounds or reduced contrast conditions, reflecting the inherent difficulty of detecting subtle early-stage symptoms in natural field environments.

The qualitative results further indicate that YOLOv10 maintains a stable detection performance across varying illumination conditions and complex canopy backgrounds. Compared to YOLOv9, YOLOv10 produces tighter bounding boxes with fewer localization errors, suggesting improved spatial consistency. However, slight variations in confidence scores are still observed when detecting low-contrast early symptoms, indicating that subtle visual features remain challenging under natural field conditions.

The qualitative detection results of YOLOv10 are shown in Figures 23 and 24. Figure 23 illustrates detection of a healthy oil palm leaf, while Figure 24 presents the localization of early Ganoderma infection symptoms. Bounding boxes and confidence scores are overlaid on each frame to support visual interpretation of model behavior under realistic field conditions.

Figure 23 shows the detection result of a healthy oil palm leaf using YOLOv10. The model correctly classified the leaf as Healthy with a confidence score of 0.96.

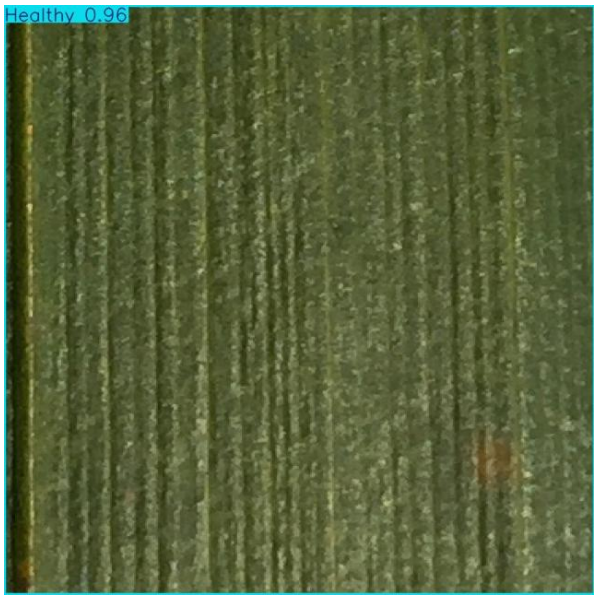


Figure 23. YOLOv10 Healthy Leaf Detection

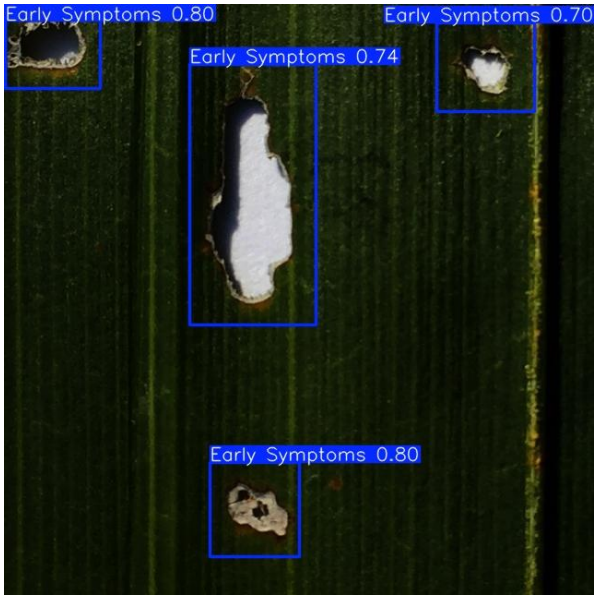


Figure 24. Early Detection of YOLOv10 Infection

Figure 24 shows the early detection result of Ganoderma infection using YOLOv10. The model identified multiple infected regions with confidence scores ranging from 0.70 to 0.80. The bounding boxes remained aligned with the lesion areas.

YOLOv11 demonstrated competitive qualitative detection performance during video-based testing, with well-aligned bounding boxes and consistently high confidence scores across detected instances. The model maintained stable localization under natural field illumination and complex canopy backgrounds, with minimal observable boundary deviations across consecutive frames.

Figure 25 and Figure 26 illustrate representative detection outputs for the healthy and early-infected classes, respectively. The overlaid bounding boxes and confidence scores demonstrate the model's ability to localize target regions consistently across frames captured under varying natural field illumination conditions.

The qualitative detection results of YOLOv11 are presented in Figure 25 and Figure 26.



Figure 25. YOLOv11 Healthy Leaf Detection

Figure 25 shows the detection result of a healthy oil palm leaf using YOLOv11. The model correctly classified the leaf as Healthy with a confidence score of 0.99.



Figure 26. Early Detection of YOLOv11 Infection

Figure 26 shows the early detection result of Ganoderma infection using YOLOv11. The model detected multiple infected regions with confidence scores ranging from 0.67 to 0.77. Although the

confidence values were lower than those of healthy leaves, the bounding boxes remained consistent with the lesion areas.

In addition, YOLOv11 demonstrates superior temporal consistency during video-based inference, with minimal fluctuations in detection confidence across consecutive frames. This stability is particularly important for real-time monitoring systems, as it reduces false alarms and improves reliability in continuous observation scenarios. Despite slightly lower confidence in detecting subtle symptoms compared to more prominent cases, the model maintains robust localization performance.

3.2 Discussion

This section discusses the experimental findings in relation to model performance, detection challenges, and practical implications for real-world deployment. The analysis focuses on the comparative performance of YOLOv9, YOLOv10, and YOLOv11 in detecting early symptoms of Ganoderma boninense under natural field conditions, as well as the trade-offs between detection accuracy and computational efficiency.

The progression from YOLOv9 to YOLOv11 reflects continuous improvements in balancing detection accuracy and computational efficiency. YOLOv9 achieved the highest accuracy, demonstrating strong capability in capturing subtle disease features. In contrast, YOLOv11 provided the fastest inference speed (35 FPS; 28.5 ms), indicating better suitability for real-time deployment. These results highlight a trade-off where YOLOv9 is more appropriate for accuracy-oriented applications, while YOLOv11 offers a more practical solution for real-time monitoring systems.

Consistent patterns across all architectures showed lower performance for the early symptoms class (precision: 0.75–0.81) than for the healthy class (precision: 0.89-0.93). This performance gap reflects inherent challenges in detecting subtle visual manifestations including mild chlorosis, slight discoloration, and early wilting that overlap with symptoms of abiotic stress. Confusion matrix analysis indicates that a substantial proportion of false negatives in the early symptoms category occurred in samples exhibiting low severity levels, where visual changes were minimal and required expert-level interpretation for reliable identification [17]. A similar phenomenon was reported in a study using a smartphone application for BSR detection, where MobileNetV2 achieved an accuracy of 77% with a special challenge in suboptimal lighting conditions [18]. This indicates that early symptoms detection remains limited by subtle color-texture variations that may not be sufficiently captured by RGB-based features alone.

This subsection contextualizes the performance of the YOLO-based detection system against representative alternative approaches, covering both laboratory-based methods and remote sensing techniques previously applied in oil palm disease detection.

A comparative analysis of alternative Ganoderma detection approaches highlights the practical considerations associated with different diagnostic strategies. Laboratory-based methods such as EDTA colorimetric assays and PCR analysis offer high diagnostic sensitivity; however, they require destructive sampling, laboratory processing time (typically several days), and per-sample operational costs [19].

Hyperspectral imaging-based remote sensing techniques have also demonstrated high classification accuracy in controlled experimental settings. Nevertheless, these systems generally rely on specialized sensors and complex data processing pipelines, which may limit scalability and large-scale field deployment in commercial plantation environments [20].

In contrast, the YOLOv11-based system developed in this study operates using standard RGB imagery that can be acquired with widely available imaging devices such as smartphones or consumer-grade UAV platforms. Despite relying on accessible hardware, the proposed system achieved an mAP50 of 0.985 and an F1-score of 0.955 on the independent testing set. Previous UAV-based studies utilizing modified infrared sensors reported lower overall classification accuracy (72.73%), suggesting that optimized RGB-based deep learning approaches can provide a practical and scalable alternative under typical field conditions [21].

This subsection evaluates the computational performance of the three YOLO architectures in terms of inference speed and processing efficiency, with reference to deployment requirements in real-time agricultural monitoring systems.

Inference performance evaluation shows that YOLOv11 achieves an optimal balance between accuracy and processing speed with an average inference time of 28.5 ms per frame (35 FPS), outperforming YOLOv10 ($32.1\ ms$, 31 FPS) and YOLOv9 ($36.8\ ms$, 27 FPS). This speed meets the requirements for real-time monitoring on commercial farms. YOLOv8 deployment studies on edge devices such as Raspberry Pi 4B with Intel Neural Compute Stick achieved an inference time of 0.4 seconds per image (2.4 FPS) with a low power consumption of $1.6W$, demonstrating the feasibility of deploying resource-constrained devices for precision agriculture applications [22].

Visualization of the detection results demonstrates high temporal consistency, with minimal frame-to-frame fluctuations in confidence scores (standard deviation $<$ 0.08), indicating predictive stability that is essential for operational decision-making. The model maintained stable localization performance under visually complex field conditions, including partial occlusion and varying illumination. However, a stratified evaluation

across controlled lighting scenarios was not conducted in this study [23].

The implementation of the YOLOv11-based detection system offers a practical framework for supporting large-scale monitoring of oil palm plantations. With a recall value of 0.971, the model demonstrates strong capability in identifying early infection cases, potentially enabling earlier intervention compared to conventional visual inspection. Such early detection may contribute to mitigating disease progression and supporting more timely management decisions [24].

The system can be integrated with unmanned aerial vehicles (UAVs) or ground-based imaging platforms to facilitate automated surveillance in plantation environments. Compared to manual inspection methods, automated detection systems may improve monitoring efficiency and consistency while reducing operator subjectivity. However, the present study did not conduct a detailed operational cost analysis. Therefore, future research should include a comprehensive economic feasibility assessment incorporating equipment costs, deployment scalability, yield recovery estimation, and long-term maintenance considerations to validate practical implementation viability.

Despite promising results, several limitations must be acknowledged. First, the dataset was collected from a single geographical location under relatively consistent environmental and lighting conditions, which may limit generalizability to different agroecological zones. Second, although overall detection performance was high, the remaining false negatives in the early symptoms class indicate ongoing challenges in identifying extremely subtle visual manifestations. Additionally, since the dataset was extracted from continuous video frames, potential temporal redundancy may exist despite augmentation strategies. Future research should focus on multi-regional dataset expansion, integration of multispectral or thermal imaging modalities to enhance pre-symptomatic detection sensitivity, and large-scale field validation under diverse plantation conditions. Exploration of lightweight model deployment on edge devices and the application of federated learning approaches may further improve system adaptability and scalability prior to widespread commercial adoption [25].

4. Conclusions

In conclusion, this study developed an automated early detection system for Ganoderma boninense infection in oil palm plantations using YOLO-based architectures. Comparative evaluation demonstrates that YOLOv9 achieved the highest detection accuracy, with an mAP50 of 0.989, recall of 0.977, F1-score of 0.968, and mAP50–95 of 0.840. In contrast, YOLOv11 delivered superior real-time performance, achieving 35 FPS with an inference time of 28.5 ms per frame. These results highlight a clear trade-off between detection precision and computational efficiency. While YOLOv9 is more suitable for applications requiring maximum detection robustness, YOLOv11 provides a more practical balance between accuracy and processing speed for real-time deployment in precision agriculture systems. Future research should focus on multi-regional dataset expansion, multispectral integration to enhance pre-symptomatic detection sensitivity, and large-scale field validation to strengthen generalizability and operational reliability. This study contributes a structured evaluation framework that integrates accuracy metrics, threshold-dependent F1 analysis, confusion matrix interpretation, and real-time inference benchmarking for agricultural disease detection.

Acknowledgements

The authors would like to express their sincere gratitude to Institut Teknologi Sawit Indonesia for providing institutional support and research facilities that made this study possible. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declare no conflict of interest.

References

[1] M. Al-Husaini, A. R. Raharja, V. Hafizh, C. Putra, and H. Hen, "Enhanced Plant Disease Detection Using Computer Vision YOLOv11," *Journal of Artificial Intelligence Research*, vol. 7, no. 1, pp. 82–95, 2025, doi: https://doi.org/10.47709/cnahpc.v7i1.5146.

[2] S. Noon and others, "Oil palm tree detection in UAV imagery using an enhanced RetinaNet," *Comput. Electron. Agric.*, vol. 227, p. 109530, 2024, doi: 10.1016/j.compag.2024.109530.

[3] M. Ilham, "Anatomical Structure of Young Oil Palm Leaves for Callogenesis Initiation," *BioMedich*, vol. 14, no. 2, pp. 763–769, 2025, doi: 10.14421/biomedich.2025.142.763-769.

[4] J. Kurihara, V. Koo, C. W. Guey, and Y. P. Lee, "Early Detection of Basal Stem Rot Disease in Oil Palm Tree Using Unmanned Aerial Vehicle-Based Hyperspectral Imaging," *Remote Sens. (Basel).*, vol. 14, no. 3, p. 799, 2022, doi: 10.3390/rs14030799.

[5] D. A. S. Daranagama and W. Takeuchi, "Smartphone application for detecting and visualizing basal stem rot disease stages in oil palm," *Frontiers in Remote Sensing*, 2025, doi: 10.3389/frsen.2025.1553844.

[6] Y. H. Haw and others, "Classification of basal stem rot using deep learning: a review of digital data collection and palm disease classification methods," *PeerJ Comput. Sci.*, pp. 1–30, 2023, doi: 10.7717/peerj-cs.1325.

[7] P. Ahmadi, S. Mansor, B. Farjad, and E. Ghaderpour, "Unmanned Aerial Vehicle (UAV)-Based Remote Sensing for Early-Stage Detection of Ganoderma," *Remote Sens. (Basel).*, vol. 14, no. 5, p. 1239, 2022, doi: 10.3390/rs14051239.

[8] A. Minarto, M. H. Ramadhan, A. Lie, and E. Irwansyah, "Oil palm tree counting and abnormality assessment using deep learning and vegetation index," *Procedia Comput. Sci.*, vol. 245, pp. 768–777, 2024, doi: 10.1016/j.procs.2024.10.303.

[9] P. Ahmadi, H. A. Araji, and B. Lu, "Convolutional SVM Networks for Detection of Ganoderma boninense at Early Stage in Oil Palm Using UAV and Multispectral Pleiades Images," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2023, pp. 25–30. doi: 10.5194/isprs-annals-X-4-W1-2022-25-2023.

[10] E. G. Jacq and Y. D. S. Marcelian, "Identification and Percentage of Disease Pathogen Attacks on Primary Palm Oil Crops," *J. Trop. Agric.*, pp. 243–255, 2023, doi: https://doi.org/10.47709/joa.v2i03.3248.

[11] I. Dwita and S. Tarigan, "Evaluation of Machine Learning Classification Algorithm for the Final Score Category of Employee Performance Allowance," *Journal of Information*

Systems, vol. 6, no. 3, pp. 251–261, 2023, doi: 10.37396/jsc.v6i3.246

[12] M. Farhan, M. Nishat, A. Elmi, and A. Bakar, "Efficient real-time palm oil tree detection and counting using YOLOv8 deployed on edge devices," *Journal of Umm Al-Qura University Engineering and Architecture*, vol. 16, no. 4, pp. 1293–1308, 2025, doi: 10.1007/s43995-025-00164-7.

[13] M. I. Anuar, H. Arof, N. M. Nor, Z. Hashim, and I. A. Seman, "Remote Sensing for Detection of Ganoderma Disease and Bagworm Infestation in Oil Palm," *Advances in Agricultural and Food Research Journal*, vol. 2, no. 1, pp. 1–14, 2021, doi: 10.36877/aafrj.a0000189.

[14] C. Mirande-Ney, G. Tcherkez, T. Balliau, M. Zivy, and F. Gilard, "Metabolic leaf responses to potassium availability in oil palm (Elaeis guineensis Jacq.) trees grown in the field," *Environ. Exp. Bot.*, vol. 175, p. 104062, 2020, doi: 10.1016/j.envexpbot.2020.104062.

[15] C. Augustine, T. H. Tee, and Y. X. Teoh, "Discovering the Ganoderma boninense Detection Methods Using Machine Learning," *IEEE Access*, vol. 9, pp. 105776–105787, 2021, doi: 10.1109/ACCESS.2021.3098307.

[16] A. Y. Putri and R. Widyowati, "Pharmacological Activity from Various Parts of Palm Oil (Elaeis guineensis Jacq.): A Review," *Journal of Pharmacognosy*, vol. 10, no. 2, pp. 143–160, 2025, doi: https://doi.org/10.21776/ub.pji.2025.010.02.9.

[17] K. H. Heng, A. A. Malik, M. Azam, B. Osman, Y. Yusop, and I. H. Hamzah, "Healthy and Unhealthy Oil Palm Tree Detection Using Deep Learning Method," *Journal of Engineering Science and Technology*, vol. 16, no. 4, pp. 748–757, 2025, doi: http://dx.doi.org/10.14569/IJACSA.2025.0160474.

[18] N. A. Husin, M. Hamim, A. Aziz, and S. Khairunniza-Bejo, "Detection and Classification of Basal Stem Rot Disease in Oil Palm Using Machine Learning Techniques: A Mini Review," *Advances in Agricultural and Food Research Journal*, 2023, doi: 10.36877/aafrj.a0000365.

[19] A. Thapa, T. Horanont, B. Neupane, and J. Klaylee, "Classification of oil palm tree conditions from UAV imagery using the YOLO object detector," *Big Earth Data*, pp. 1–27, 2025, doi: 10.1080/20964471.2025.2491881.

[20] N. A. Husin, "Ganoderma BSR Disease Detection Using Machine Learning with UAV Imagery Oil Palm Features, Thermal and Visible Spectral Indices," 2021. doi: 10.2139/ssrn.5304086.

[21] Rista, "Middle-Level Fusion YOLO on Multispectral Image to Detect Unhealthy Oil Palm Trees," in *Journal of Physics: Conference Series*, IOP Publishing, 2024, p. 12045. doi: 10.1088/1742-6596/2866/1/012045.

[22] L. Palupi, E. Ihsanto, and F. Nugroho, "Analysis of Validation and Evaluation of Ginger Variant Object Detection Model Using YOLOv5 Algorithm," *Journal of Soft Computing and Artificial Intelligence*, vol. 5, no. 1, 2023, doi: 10.47065/josh.v5i1.4380.

[23] M. Wahyuni, T. Sabrina, and H. Santoso, "Using Machine Learning in Detecting Ganoderma Disease in Oil Palm Plants," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, pp. 145–155, 2024, doi: https://www.ijisae.org/index.php/IJISAE/article/view/4801.

[24] F. Rokhana and others, "Defense response changes in roots of oil palm (Elaeis guineensis Jacq.) seedlings after internal symptoms of Ganoderma boninense," *BMC Plant Biol.*, vol. 22, pp. 1–23, 2022, doi: 10.1186/s12870-022-03493-0.

[25] D. A. Saragih, I. Ovie, Y. Sitompul, D. R. Pulungan, and R. Rambe, "Factors Analysis on Palm Oil (Elaeis guineensis Jacq.) Production," *International Journal of Science and Management Studies*, vol. 1, pp. 2192–2200, 2023, doi: https://doi.org/10.30596/miceb.v1i0.330