ISSN : 2355-9365 e-Proceeding of Engineering : Vol.12, No.4 Agustus 2025 | Page 5130

YOLOv11 for Classification of Strawberry Quality and Ripeness

1st Ardy Fahriansyah
School of Electrical Engineering
Telkom University
Bandung, Indonesia
ORCID : 0009-0001-7594-997X

2nd Pavel Manaf El Zaky
School of Electrical Engineering
Telkom University
Bandung, Indonesia
ORCID: 0009-0000-1903-6618

3rd Ledya Novamizanti
School of Electrical Engineering
CoE of AILO
Telkom University
Bandung, Indonesia
ledyaldn@telkomuniversity.ac.id

4th Sofia Sa'idah
School of Electrical Engineering
Telkom University
Bandung, Indonesia
sofiasaidahsfi@telkomuniversity.ac.id

Strawberries hold significant economic value in Indonesia due to their high demand and nutritional benefits. Traditional harvesting methods, which rely on manual visual inspection, are often inefficient and prone to errors. Real-time multi-object detection presents a promising solution to enhance automation in harvesting, ripeness classification, and post-harvest processing. This study assesses the performance of four YOLOv11 variants—YOLOv11N, YOLOv11S, YOLOv11M, and YOLOv11L—in detecting strawberries across five quality and ripeness categories: Unripe, Half Ripe Grade B, Half Ripe Grade A, Fully Ripe Grade B, and Fully Ripe Grade A. A dataset originally consisting of 3,055 high-resolution strawberry images was expanded through data augmentation to 7,940 images. These were subsequently split into training (7,330 images), validation (305 images), and testing (305 images) sets. All models were trained under identical conditions utilizing the AdamW optimizer, cosine annealing learning rate scheduling, a batch size of 16, and an input resolution of 640×640 pixels. Performance was evaluated based on Precision, Recall, F1-Score, mAP@0.5, mAP@0.95, and inference time. The results indicate that YOLOv11N achieved the best overall performance, with a Precision of 0.869, Recall of 0.878, F1-Score of 0.87, mAP@0.95 of 0.830, and the fastest inference time of 3.6 ms, rendering it suitable for real-time deployment. YOLOv11M provided a balanced trade-off between accuracy and speed, while YOLOv11S offered competitive accuracy with lower inference latency. YOLOv11L demonstrated strong detection capabilities but with the slowest inference time. These findings affirm the efficacy of YOLOv11-based models in facilitating scalable and intelligent systems for precision agriculture.

Keywords—strawberry, computer vision, deep learning, object detection, classification, YOLO.

I. INTRODUCTION

The adoption of artificial intelligence (AI) and deep learning technologies has led to a significant transformation in modern agriculture, enabling automation that improves productivity and decision-making [1]. One notable application of AI in this domain is real-time fruit detection, which enhances harvest efficiency and reduces post-harvest losses [2]. Manual fruit harvesting remains labor-intensive and error-prone, making it less effective for large-scale agricultural operations [3]. Inconsistent human performance further complicates manual inspection, emphasizing the need for automation to ensure quality and yield optimization [4].

Strawberries are particularly difficult to detect due to their soft texture, variable sizes, and multiple ripeness stages [5]. Harvesting strawberries at incorrect maturity levels can lead to reduced shelf life and financial losses [6]. The tendency of strawberries to grow in clusters contributes to frequent occlusions that hinder accurate detection [7]. Environmental conditions such as lighting variability and complex backgrounds further reduce model reliability in open-field settings [8].

Recent advances in deep learning, especially convolutional neural networks (CNNs), have shown high effectiveness in object detection for agriculture [9]. Among these, the YOLO (You Only Look Once) framework has become a leading solution due to its balance of speed and accuracy [10]. Earlier versions like YOLOv3 and YOLOv4 have been applied in fruit detection but often struggle with complex field conditions [11]. Enhanced versions such as YOLOv7 and YOLOv11 offer architectural improvements that improve both detection precision and computational efficiency [12].

Nevertheless, challenges remain in deploying these models for real-time multi-object strawberry detection due to occlusions, varied lighting, and intra-class variability [13]. To address this, a comprehensive evaluation of YOLOv11 variants is needed to understand their performance in practical agricultural environments [14]. This study investigates four YOLOv11 variants, YOLOv11L, YOLOv11M, YOLOv11S, and YOLOv11N using key performance metrics including precision, recall, F1-score, mAP (0.5 and 0.95), and inference time [15].

The structure of this paper is organized as follows. Section II reviews related work relevant to the study. Section III describes the proposed method, including dataset preparation, model training procedures, and evaluation metrics. Section IV presents the experimental results along with a detailed performance analysis. Finally, Section V concludes the paper by summarizing key findings and outlining potential directions for future research.

II. RELATED WORK

Recent developments in deep learning have significantly advanced the application of YOLO-based models in fruit detection for precision agriculture. Azizah et al. [6] proposed a hybrid framework combining YOLOv7 and EfficientNetV2S for strawberry quality and ripeness classification. Their model successfully categorized strawberries into five distinct classes under various natural lighting conditions, achieving precision, recall, and F1-score of 0.99. However, the system required a two-stage process involving local inference (180 ms) and cloud-based post-processing (1–2 s), making it less suitable for real-time edge deployment. Akhyar et al. [16] introduced Lightning YOLOv4, a lightweight model optimized for detecting surface defects in sawn lumber, which maintained high accuracy and low computational cost for real-time inspection in industrial settings. Aziz et al. [17] evaluated several YOLOv8 models for skipjack fish quality assessment, showing strong accuracy (mAP 93–95%) but moderate inference speeds (20–30 ms), emphasizing the need to match architecture to application constraints. Similarly, Akhyar et al. [18] compared YOLOv5, v6, and v7 in lobster surveillance, focusing on behavioral detection under controlled aquaculture conditions. Wu et al. [19] enhanced YOLOv5 with CBAM to detect pomegranates in complex orchard scenes, achieving 92.2% mAP with a 115 ms inference speed, highlighting improvements in small and occluded object detection.

In comparison to these prior works, the YOLOv11 models proposed in this study demonstrated competitive or superior performance across various metrics. The YOLOv11N variant achieved an mAP@0.5 of 94.4% and F1-score of 0.87, with an extremely low inference time of 3.6 ms, enabling real-time edge deployment. YOLOv11M also performed well with mAP up to 96.4% and inference time under 6 ms. These results outperform or match prior models in accuracy while offering significantly faster detection speeds—up to 30× faster than attention-based models such as YOLO-Granada. Moreover, unlike Lightning YOLOv4 [16], which was limited to binary classification, YOLOv11 supports fine-grained multi-class ripeness grading under real-field conditions. These comparisons confirm that YOLOv11N and M deliver high accuracy and superior speed, making them well-suited for real-time fruit quality detection in agricultural automation.

Beyond data augmentation, various methods have been proposed to tackle data scarcity in deep learning. Alzubaidi et al. [21] reviewed strategies like transfer learning, self-supervised learning, GANs, and DeepSMOTE. While this study focuses on augmentation, future work could explore these approaches to enhance performance under limited data. These studies also emphasize YOLO's adaptability across diverse agricultural contexts.

These studies collectively highlight the adaptability of YOLO-based architectures across diverse agricultural scenarios. However, comprehensive evaluations of newer YOLO variants, such as YOLOv11, in multi-class ripeness classification under field conditions were previously lacking—this study addresses that gap.

III. Proposed Method

A. Dataset

The dataset used in this study consists of 3,055 high-resolution strawberry images collected from both commercial farms and controlled environments to ensure variability. Most images were captured at Ichigo Farm in Ciwidey, Indonesia, using consumer-grade smartphones such as the Redmi Note 9 Pro, Realme 9 Pro+, and Samsung S21 under diverse natural lighting conditions. Each image was annotated with bounding boxes to identify strawberries at various ripeness and quality levels. The dataset is categorized into five classes—UNR, HRB, HRA, FRB, and FRA—with 611 images per class. As shown in Fig. 1 and Table I, these labels provide a structured basis for learning. The data is evenly divided into training (489), validation (61), and testing (61) sets per class, enabling balanced model evaluation across categories.

After applying data augmentation techniques—including horizontal and vertical flipping, 90-degree rotations, and random rotations between −15° and +15°—the dataset expanded to 7,940 images, comprising 7,330 for training and 305 each for validation and testing. The dataset is categorized into five ripeness-quality classes: Unripe (UNR), Half Ripe B (HRB), Half Ripe A (HRA), Fully Ripe B (FRB), and Fully Ripe A (FRA), with each class containing 1,588 images. Of these, 1,466 images per class are used for training and 61 each for validation and testing. This consistent allocation and labeling strategy ensures a balanced and comprehensive dataset for training deep learning models in strawberry quality and ripeness classification.

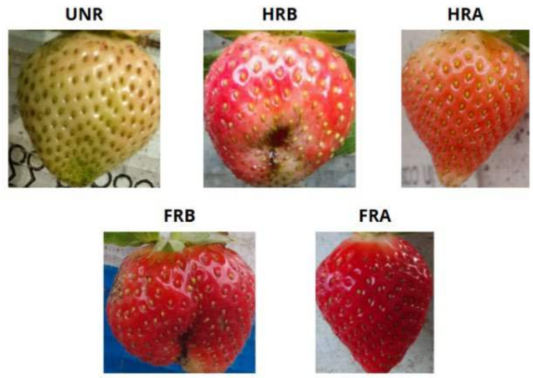


Figure 1
The strawberry samples are divided into five categories, including Unripe (UNR), Half Ripe – Type B (HRB), Half Ripe – Grade A (HRA), Fully Ripe – Grade B (FRB), and Fully Ripe – Grade A (FRA) (from left to right).

TABLE I.
Summary of Strawberry Ripeness Classes and Quality Descriptions

Class	Ripeness Category	Description
UNR	Unripe	Green color with minimal red; firm texture; not suitable for harvest.
HRB	Half Ripe (Grade B)	Partially red with uneven texture; visible seeds; lower visual quality; early ripening stage.
HRA	Half Ripe (Grade A)	More even red coloration; smoother surface; less visible seeds; better texture and visual appeal.
FRB	Full Ripe (Grade B)	Deep red color with minor surface imperfections; slightly less uniform than premium grade.
FRA	Full Ripe (Grade A)	Uniform deep red color; smooth, high-grade texture; market-ready appearance with optimal quality.

B. YOLOV11 models and Evolution Metrics

This study evaluates YOLOv11, which represent different generations of the You Only Look Once (YOLO) object detection architecture.

1) *YOLOv11 Variants*

Fig. 2 is the framework of the proposed YOLOv11 based object detection models for strawberry quality and ripeness classification. YOLOv11 comprises four scalable model variants—Nano (N), Small (S), Medium (M), and Large (L)—each defined by three architectural parameters: depth_multiple, width_multiple, and max_channels. These parameters adjust the model's complexity by scaling the number of layers and channels, enabling a trade-off between detection accuracy and computational efficiency.

YOLOv11N is the most lightweight variant with a depth_multiple of 0.50 and width_multiple of 0.25, using up to 1024 channels. It contains 181 layers, 2.6 million parameters, and operates at 6.6 GFLOPs, optimized for edge devices and real-time applications. YOLOv11S doubles the width (0.50) while retaining the same depth, resulting in 9.4 million parameters and 21.7 GFLOPs—balancing speed and accuracy. YOLOv11M further increases the width to 1.00 with the same depth (0.50) and a reduced max_channels of 512, yielding 20.1 million parameters and 68.5 GFLOPs. YOLOv11L, the largest model, uses a depth and width of 1.00, resulting in 25.3 million parameters, 88.6 GFLOPs, and 357 layers, providing the highest detection precision at the cost of speed.

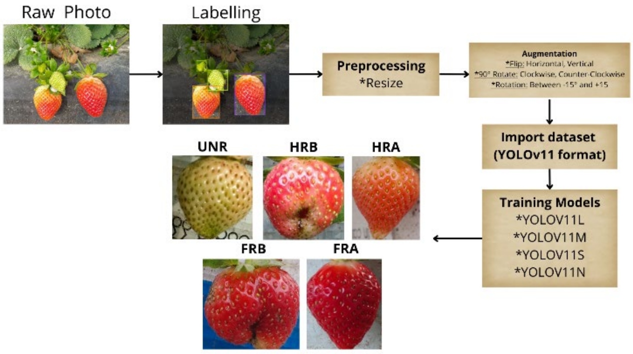


Figure 2
Framework of the proposed YOLOv11 Based Object Detection Models for Strawberry Quality and Ripeness Classification.

In terms of architecture, all YOLOv11 variants share the same fundamental structure consisting of a Backbone, Neck, and Head. The Backbone incorporates a C3K2 block, which is an enhanced version of the CSP (Cross Stage Partial) module using 3×3 convolutional kernels. This block improves feature extraction efficiency by maintaining gradient flow and reducing computational overhead. The Neck utilizes a Spatial Pyramid Pooling – Fast (SPPF) module that aggregates multi-scale contextual features, enhancing the model's ability to detect objects at different sizes and resolutions.

The Head adopts a C2PSA (Convolution + Parallel Spatial Attention) module, which introduces spatial attention mechanisms to focus on the most relevant regions of the feature map. This improves both localization and classification accuracy, particularly in complex scenes with overlapping or occluded objects. The combination of a lightweight backbone and attention-guided head allows YOLOv11 to maintain a balance between fast inference and high precision, making it suitable for deployment across a variety of precision agriculture and real-time detection applications.

2) *Evaluation Metrics*

To assess the performance of YOLOv11, the following key evaluation metrics were used:

a) Precision

Precision measures the proportion of correctly detected strawberries among all instances that the model predicted as strawberries. It is defined by the equation:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

with TP (True Positives) are correctly detected strawberries, and FP (False Positives) are incorrectly predicted strawberries—instances that were labeled as strawberries by the model but were not actual strawberries. A high precision value indicates that the model makes few false positive errors, which is important in applications where incorrect detection can cause unnecessary actions, such as robotic harvesting of unripe or non-existent fruits [20].

b) Recall

Recall measures the model's ability to detect all actual strawberries present in the image. It is calculated using the formula:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

Where FN (False Negatives) are missed detections—actual strawberries that were not identified by the model. A high recall means the model successfully captures most of the objects of interest, which is critical in scenarios where missing ripe strawberries may result in yield loss [20].

c) F1-score

The F1-Score is the harmonic mean of precision and recall, providing a single metric that balances the trade-off between them:

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

The F1-Score is especially useful when there is an uneven class distribution or when both false positives and false negatives carry significant consequences. A high F1-Score indicates a good balance between precision and recall, making it a reliable indicator of the model's overall detection performance in realistic agricultural conditions [17].

d) Mean Average Precision (mAP)

The mean Average Precision (mAP) is a standard metric for evaluating object detection accuracy, particularly in terms of bounding box localization. mAP@0.5 measures precision at a single IoU threshold of 0.5, while mAP@0.95 averages precision across multiple IoU thresholds (0.5 to 0.95), offering a more comprehensive assessment of detection performance.

IV. EXPERIMENTS AND RESULTS

This section presents the experimental results and performance analysis of YOLOv11 in real-time multi-object strawberry detection. The models were evaluated based on Precision, Recall, F1-Score, mean Average Precision (mAP 0.5 & mAP 0.95), and inference time. Additionally, qualitative analysis using test images is provided to visualize detection accuracy.

A. *Training Configuration*

To ensure an objective evaluation, all YOLOv11 model variants were trained under uniform experimental conditions. This consistent setup allows for a fair comparison of performance across different model architectures and ensures the reproducibility of results. The training parameters of the proposed model are presented in Table II.

B. *Training Performance Analysis*

This section compares the training performance of YOLOv11N, YOLOv11S, YOLOv11M, and YOLOv11L using precision, recall, F1-score, mAP@0.5, mAP@0.95, and inference time. Visual analyses, including confusion matrices and per-class performance, further evaluate their effectiveness in detecting fruit ripeness and quality.

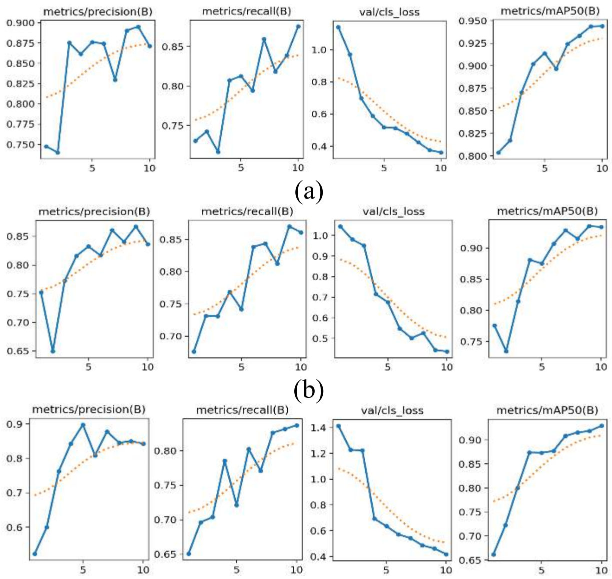
TABLE II.
TRAINING PARAMETERS OF THE PROPOSED MODEL

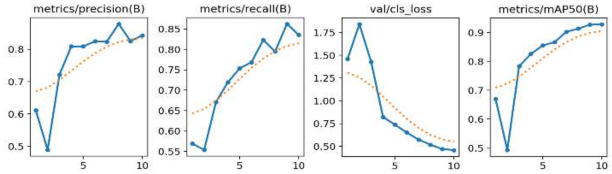
Component	Description
Image Processing	Resized to 640×640 pixels, normalized values
Augmentation	Flip: Horizontal, Vertical, 90° Rotate: Clockwise, Counter-Clockwise, Rotation: Between -15° and +15°
Batch Size	16
Epochs	10
Optimizer	AdamW (momentum=0.9)
Learning Rate	0.001111
Training Hardware	Google Colab GPU
platform utilized	NVIDIA Tesla T4 GPU with 16 GB of VRAM and 25 GB of available RAM

All YOLOv11 variants were trained using Google Colab, supported by an NVIDIA Tesla T4 GPU with 16 GB VRAM and 25 GB RAM. This configuration provided sufficient computing power for processing the augmented dataset efficiently using a batch size of 16 and an input resolution of 640×640 pixels. The use of a cloud-based platform ensured accessibility and reproducibility of the experiments while maintaining consistent conditions across all model variants. This setup also reflects a practical and cost-effective environment for deep learning applications in agriculture.

C. *Training Performance Analysis*

This section compares the training performance of YOLOv11N, YOLOv11S, YOLOv11M, and YOLOv11L using precision, recall, F1-score, mAP@0.5, mAP@0.95, and inference time. Visual analyses, including confusion matrices and per-class performance, further evaluate their effectiveness in detecting fruit ripeness and quality.





(d)
Figure 3
Performance training model results (a) YOLOV11N, (b) YOLOV11S, (c) YOLOV11M, (d) YOLOV11L

Fig. 3 and Tables III to VI collectively illustrate the comparative performance of the four YOLOv11 variants—Nano (N), Small (S), Medium (M), and Large (L)—across five strawberry ripeness classes. Fig. 3 highlights YOLOv11N and YOLOv11M as the top performers in terms of training efficiency and final accuracy, showcasing superior metrics such as precision, recall, classification loss, and mAP@0.5. YOLOv11N stands out with rapid convergence and stable performance, making it ideal for low-latency, high-accuracy applications, while YOLOv11M offers a well-balanced trade-off between speed and precision for real-world deployment. In contrast, YOLOv11S and YOLOv11L, though still viable, showed slightly less training stability or efficiency. Detailed evaluation in Tables III–VI further supports these observations: YOLOv11L achieved a precision of 0.843, recall of 0.835, and mAP@0.5 of 0.928, with FRB and HRB leading in precision (0.890) and recall (0.924), respectively. YOLOv11M posted nearly identical results, while YOLOv11S yielded a higher mAP@0.5 of 0.934 and excellent recall (0.861), despite HRB having the lowest precision (0.767) but highest recall (0.970). These findings demonstrate the nuanced strengths of each variant and affirm the effectiveness of YOLOv11 for precise strawberry ripeness classification.

TABLE III.
PERFORMANCE OFYOLOV11N

Class	Precision	Recall	Map50	Map50-95
UNR	0.886	0.772	0.898	0.677
HRB	0.784	0.936	0.958	0.865
HRA	0.914	0.848	0.941	0.87
FRB	0.898	0.896	0.966	0.887
FRA	0.864	0.939	0.956	0.853
Average	**0.869**	**0.878**	**0.944**	**0.83**

TABLE IV.
PERFORMANCE OFYOLOV11S

Class	Precision	Recall	Map50	Map50-95
UNR	0.863	0.817	0.901	0.685
HRB	0.767	0.97	0.955	0.857
HRA	0.873	0.781	0.93	0.842
FRB	0.893	0.842	0.947	0.868
FRA	0.78	0.894	0.938	0.825
Average	**0.835**	**0.861**	**0.934**	**0.815**

TABLE V.
PERFORMANCE OFYOLOV11 M

Class	Precision	Recall	Map50	Map50-95
UNR	0.868	0.794	0.881	0.677
HRB	0.784	0.955	0.954	0.866
HRA	0.87	0.761	0.923	0.838
FRB	0.903	0.826	0.949	0.871
FRA	0.789	0.85	0.939	0.811
Average	**0.843**	**0.837**	**0.929**	**0.813**

TABLE VI.
PERFORMANCE OFYOLOV11 L

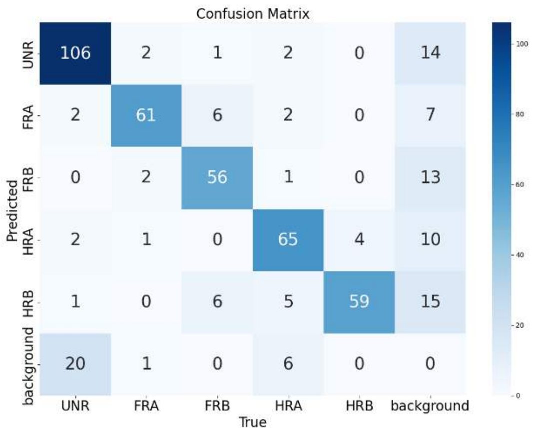
Class	Precision	Recall	Map50	Map50-95
UNR	0.862	0.809	0.908	0.684
HRB	0.822	0.924	0.959	0.852
HRA	0.818	0.759	0.91	0.829
FRB	0.89	0.819	0.93	0.836
FRA	0.825	0.864	0.935	0.823
Average	**0.843**	**0.835**	**0.928**	**0.805**

Among all variants, YOLOv11N achieved the best overall performance with the highest precision (0.869), recall (0.878), and mAP@0.5 (0.944), along with the fastest inference time (3.6 ms). It excelled in classification accuracy across multiple classes, including HRA (precision 0.914) and FRA (recall 0.939), although slightly lower results were noted for the UNR class. These findings highlight the distinct strengths and trade-offs of each model: YOLOv11N is most suitable for real-time applications due to its speed and accuracy; YOLOv11M offers a balanced solution for large-scale deployment; YOLOv11S provides efficient performance for semi-real-time environments; and YOLOv11L, with the slowest inference, is more appropriate for high-accuracy tasks where latency is less critical. Each variant thus serves different operational requirements within precision agriculture systems.

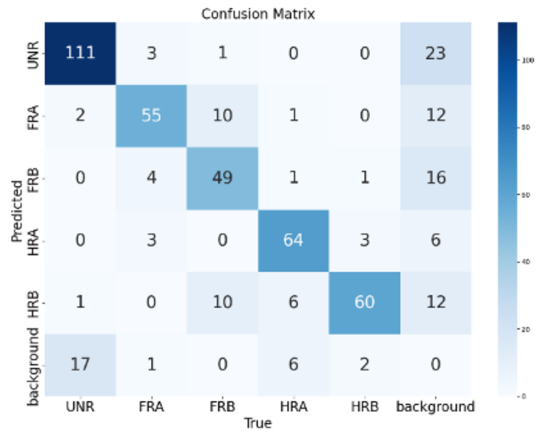
TABLE VII.
COMPARISON OF YOLOV11 VARIANT MODELS

Model	Precision (%)	Recall (%)	F1-Score	Map 0.5	Map 0.95	Inference Time (ms)
YOLOV11N	0.869	0.878	0.87	0.944	0.83	3.6
YOLOV11S	0.835	0.861	0.84	0.934	0.815	4.8
YOLOV11M	0.843	0.837	0.84	0.929	0.813	11.3
YOLOV11L	0.843	0.835	0.84	0.928	0.805	13.8

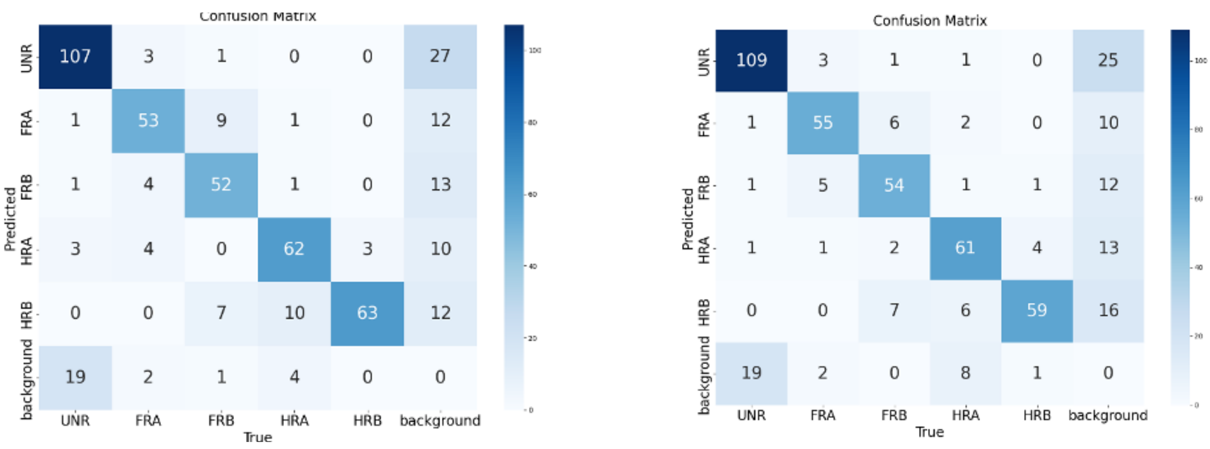
Table VII presents a comparative evaluation of four YOLOv11 variants—YOLOv11N, YOLOv11S, YOLOv11M, and YOLOv11L—based on key performance metrics: Precision, Recall, F1-Score, mAP@0.5, mAP@0.95, and Inference Time. YOLOv11N achieves the highest precision (86.9%) and recall (87.8%), along with the best mAP@0.5 (94.4%), while also delivering the fastest inference time (3.6 ms), indicating strong performance in both accuracy and efficiency. YOLOv11M and YOLOv11L show similar F1-scores but trade off speed for slightly lower mAP scores. YOLOv11S provides a balance between speed and accuracy, making it suitable for lightweight deployments.



(a)



(b)



(c) (d)

Figure 4
Confusion matrix training results in comparison models: (a) YOLOV11N, (b) YOLOV11S, (c) YOLOV11M, (d) YOLOV11L

D. Testing Performance Analysis

Fig. 4(a) presents the confusion matrix of the YOLOv11N model, which achieves the highest overall accuracy with minimal misclassification, making it the most accurate among the variants. Fig. 4(b) shows the performance of YOLOv11S, which demonstrates acceptable results but exhibits a higher tendency to misclassify FRB samples. In Figure 4(c), YOLOv11L performs reasonably well but shows increased background confusion and inter-class errors compared to other models. Finally, Fig. 4(d) illustrates the confusion matrix of YOLOv11M, which demonstrates strong and balanced classification across all strawberry ripeness classes. It achieves high accuracy in HRA and HRB with relatively low misclassification, particularly in distinguishing between visually similar categories like FRB and FRA. Compared to other variants, YOLOv11M outperforms YOLOv11L in reducing background confusion and inter-class errors, while maintaining a favorable trade-off between recall and precision. Overall, YOLOv11M offers strong generalization and remains a robust alternative for applications requiring both accuracy and computational efficiency, making it well-suited for scalable agricultural systems.

Fig. 5 demonstrates the real-world performance differences of YOLOv11 variants on strawberry detection. YOLOv11N and YOLOv11M exhibit superior detection accuracy and classification precision, even in non-ideal conditions. YOLOv11N stands out for its speed and consistent labeling, while YOLOv11M excels in reliability and class differentiation. Conversely, YOLOv11S and YOLOv11L show slightly lower detection robustness, suggesting they may be more suitable for specific controlled environments or scenarios where either speed (YOLOv11S) or deep feature extraction (YOLOv11L) is prioritized.

In comparison to previous studies in strawberry and fruit detection, the YOLOv11 models proposed in this study demonstrated competitive or superior performance across various metrics. Azizah et al. [18] implemented a YOLOv7-EfficientNetV2S hybrid model that achieved exceptional precision, recall, and F1-Score values of 99% each; however, their detection time was notably longer at 2.392 seconds per image. In contrast, the YOLOv11N model in our study delivered significantly faster inference time (3.6 ms) while maintaining high detection accuracy (mAP@0.5 of 94.4%, F1-score of 0.87), making it more viable for real-time applications.

(a) (b)

(c) (d)

Figure 5
Detection testing results of strawberry model (a) YOLOV11N, (b) YOLOV11S, (c) YOLOV11M, (d) YOLOV11L

V. CONCLUSION

This study presents a comparative evaluation of four YOLOv11 model variants YOLOv11N, YOLOv11S, YOLOv11M, and YOLOv11L for multi-class strawberry quality and ripeness detection. The results highlight trade-offs between detection accuracy and inference speed, which are critical considerations for practical deployment in precision agriculture. YOLOv11N demonstrated the best overall performance, achieving the highest precision (86.9%), recall (87.8%), mAP@0.5 (94.4%), mAP@0.95 (83.0%), and the fastest inference time (3.6 ms), making it the most suitable model for real-time applications such as robotic harvesting and edge-based monitoring. YOLOv11M offered a balanced trade-off between accuracy and processing speed, achieving consistent detection performance (mAP@0.5: 92.9%, mAP@0.95: 81.3%) with moderate inference latency (11.3 ms), making it ideal for large-scale automated sorting and grading systems. YOLOv11S provided competitive accuracy (mAP@0.5: 93.4%, mAP@0.95: 81.5%) and relatively fast inference (4.8 ms), making it a practical option for semi-real-time applications like greenhouse monitoring. Meanwhile, YOLOv11L achieved robust detection accuracy in complex scenes (mAP@0.5: 92.8%, mAP@0.95: 80.5%) but exhibited the slowest inference speed (13.8 ms), thus better suited for post-harvest quality analysis where latency is less critical.

These findings demonstrate the effectiveness of YOLOv11-based models in supporting AI-driven precision agriculture, enabling accurate, efficient, and scalable fruit classification. Future work will focus on optimizing model inference for low-power edge devices, enhancing robustness under variable field conditions, and extending evaluation to other crop types and agricultural environments.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Telkom University for the internal research grant No. 340/LIT06/PPM-LIT/2024 and to Ichigo Farm, Ciwidey, Bandung, West Java, Indonesia, for providing the strawberries dataset and sharing their expertise.

REFERENCES

[1] F. Xiao, H. Wang, Y. Xu, dan R. Zhang, "Fruit Detection and Recognition Based on Deep Learning for Automatic Harvesting: An Overview and Review," Agronomy, vol. 13, no. 6, p. 1625, Jun. 2023.

[2] H. Zhou, Y. Wang, J. Li, dan F. Chen, "Research on Kiwi Fruit Flower Recognition for Efficient Pollination Based on an Improved YOLOv5 Algorithm," Horticulturae, vol. 9, no. 3, p. 400, Mar. 2023.

[3] R. Pathak dan H. Makwana, "Classification of Fruits Using Convolutional Neural Network and Transfer Learning Models," Journal of Management Information and Decision Sciences, vol. 24, Special Issue 3, pp. 1–12, 2021.

[4] F. Wu dan J. Wang, "YOLO Deep Learning Algorithm for Object Detection in Agriculture: A Review," Journal of Agricultural Informatics, vol. 15, no. 2, pp. 34–45, 2024.

[5] X. Li dan Y. Zhang, "Comparing YOLOv11 and YOLOv8 for Instance Segmentation in Orchard Environments," arXiv preprint, arXiv:2410.19869, 2024.

[6] S. Azizah, M. Wahidin, M. Padang, L. Novamizanti, and S. Sa'idah, "Identifying the ripeness and quality level of strawberries based on YOLOv7-EfficientNet," in 2024 International Conference on Data Science and Its Applications (ICoDSA), Kuta, Bali, Indonesia, 2024, pp. 451–456, doi: 10.1109/ICoDSA62899.2024.10651988.

[7] Y. Tian et al., "Apple Detection During Different Growth Stages in Orchards Using the Improved YOLO-V3 Model," Computers and Electronics in Agriculture, vol. 157, pp. 417–426, 2019.

[8] I. Sa et al., "DeepFruits: A Fruit Detection System Using Deep Neural Networks," Sensors, vol. 16, no. 8, p. 1222, Aug. 2016.

[9] L. Fu et al., "Faster R-CNN-Based Apple Detection in Dense-Foliage Fruiting-Wall Trees Using RGB and Depth Features," Biosystems Engineering, vol. 197, pp. 245–256, 2020.

[10] M. Rahnemoonfar dan C. Sheppard, "Deep Count: Fruit Counting Based on Deep Simulated Learning," Sensors, vol. 17, no. 4, p. 905, Apr. 2017.

[11] N. Häni, P. Roy, dan V. Isler, "A Comparative Study of Fruit Detection and Counting Methods for Yield Mapping in Apple Orchards," Journal of Field Robotics, vol. 37, no. 2, pp. 263–282, 2020.

[12] A. Koirala, K. B. Walsh, Z. Wang, dan C. McCarthy, "Deep Learning – Method Overview and Review of Use for Fruit Detection and Yield Estimation," Computers and Electronics in Agriculture, vol. 162, pp. 219–234, 2019.

[13] M. Stein, S. Bargoti, dan J. Underwood, "Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry," Sensors, vol. 16, no. 11, p. 1915, Nov. 2016.

[14] S. Chen, J. Yang, dan L. Deng, "An Improved YOLOv5 Model Based on Attention Mechanism for Detection of Small-Size Fruits in Complex Backgrounds," Computers and Electronics in Agriculture, vol. 181, p. 105951, 2021.

[15] H. Kang dan C. Chen, "Apple Detection in Complex Scene Using the Improved YOLOv4 Model," Agronomy, vol. 11, no. 3, p. 476, Mar. 2021.

[16] F. Akhyar, L. Novamizanti, T. Putra, E. N. Furqon, M.-C. Chang, and C.-Y. Lin, "Lightning YOLOv4 for a Surface Defect Detection System for Sawn Lumber," in 2022 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2022, pp. 168–173, doi: 10.1109/MIPR54900.2022.00039.

[17] K. Aziz, M. D. Maulana, R. F. Adawiyah, R. F. Firdaus, L. Novamizanti, and F. Ramdhon, "Comparative analysis of YOLOv8 models in skipjack fish quality assessment system," in Proc. 2023 3rd Int. Conf. Intell. Cybern. Technol. Appl. (ICICyTA), 2023, pp. 237–242.

[18] F. Akhyar, L. Novamizanti, K. Usman, G. M. Aditya, F. N. Hakim, M. Z. Ilman, F. Ramdhon, and C.-Y. Lin, "A comparative analysis of the YOLO models for intelligent lobster surveillance camera," in Proc. 2023 Asia Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC), 2023, pp. 2131–2136.

[19] Y. Wu, H. Li, and Y. Gao, "Lightweight Attention-Enhanced YOLO Model for Pomegranate Detection in Complex Environments," Scientific Reports, vol. 14, no. 1, p. 3129, 2024.

[20] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. da Silva, "A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit," Electronics, vol. 10, no. 3, p. 279, 2021.

[21] L. Alzubaidi et al., "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications," *Journal of Big Data*, vol. 10, no. 1, p. 46, 2023. doi:10.1186/s40537-023-00727-2.