ISOMAse
International Society of Ocean, Mechanical and Aerospace Scientists and Engineers

Journal of Ocean, Mechanical and Aerospace
-Science and Engineering-
30 July 2026. Vol.70 No.2

DOI:
10.36842/jomase.v70i2.638

Performance Analysis of the YOLOv8 Algorithm for Detecting of Stacked Defective Oil Palm Fresh Fruit Bunches on a Moving Conveyor

Minarni Shiddiq[a,*], Dodi Sofyan Arief[b], Roni Salambue[c], Cindi Melinda Malau[a], Yohana Christia Navili[a], Vicky Vernando Dasta[a], Muhammad Ikhsan Hamid[a] and Nanda Syaputra[a],

[a)] *Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Riau, Indonesia*
[b)] *Department of Mechanical Engineering, Faculty of Engineering, Universitas Riau, Indonesia*
[c)] *Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Riau, Indonesia*

*Corresponding author: minarni.shiddiq@lecturer.unri.ac.id

Paper History

Received: 04-June-2026
Received in revised form: 20-July-2026
Accepted: 30-July-2026

ABSTRACT

Crude palm oil (CPO) is the leading export commodity for countries such as Indonesia and Malaysia. The quality of CPO depends on the raw material and oil palm fresh fruit bunches (FFB). Various sorting and grading methods based on computer vision and machine learning have been developed to assess FFB quality automatically. However, most research has focused on fruit ripeness. In fact, empty bunches, rotten fruit, long stalks, and thorny or spiky bunches are also sorting parameters and are categorized as defective FFBs. This study aims to evaluate the performance of the YOLOv8l-Seg and YOLOv8x-Seg models in detecting and segmenting normal and defective FFBs stacked on a moving conveyor. Stacked FFBs mean there is more than one FFB in a camera field of view (FOV), which is easily found during the real-time sorting process. The dataset consists of five classes: normal, long stalks, thorny, empty, and rotten bunches. Evaluation was conducted using the mean Average Precision (mAP), accuracy, precision, recall, and F1-score metrics. The results show that YOLOv8x-Seg obtained 93% accuracy and 95% mAP. The YOLOv8l-Seg reached 92% accuracy and 93.4% mAP. Therefore, both models have the potential to be applied in real-time automated oil palm FFB sorting and grading systems.

KEYWORDS: *Computer Vision, Deep Learning, Oil Palm, Stacked Fresh Fruit Bunches, YOLOv8 Algorithms.*

NOMENCLATURE

CPO	*Crude Palm Oil*
FFB	*Fresh Fruit Bunches*
YOLO	*You Look Only Once*
CNN	*Convolutional Neural Network*
mAP	*Mean Average Precision*
TP	*True Positive*
TN	*True Negative*
FP	*False Positive*
FN	*False Negative*

1. INTRODUCTION

Sorting and grading oil palm fresh fruit bunches (FFB) is a crucial step in the palm oil processing industry, as it plays a key role in determining the quality of the raw material that will be processed into crude palm oil (CPO) [1]. Through the grading process, the oil palm FFBs are classified based on their ripeness and physical conditions (defects), ensuring that only FFBs of the appropriate quality proceed to the next stage of CPO processing. The accuracy of this process directly impacts oil yield, the quality of the final product, and the mill's operational efficiency [2]. Therefore, the sorting and grading process is an integral part of the quality control system in palm oil mills. However, to date, most of the process is still performed manually and traditionally by field operators based on visual inspection of fruit colors, the number of loose fruits, and the physical conditions of the bunches. This method has various limitations that can affect the consistency of grading results, such as subjectivity, work fatigue, and lower efficiency in time and operational cost. Consequently, the FFBs that do not meet quality standards or have certain physical defects may slip through the sorting process and enter the processing stage [3].

Imaging techniques have been used to develop automatic, real-time sorting and grading of oil palm FFBs over the last few decades, especially computer vision systems. Advances in computer vision systems have opened new opportunities for developing automated sorting and grading systems in the agricultural industry [4]. Computer vision is an imaging technique that enables computers to acquire, process, and analyze images and videos to extract useful information and make decisions. Computer vision has applications in various agricultural fields because it operates more quickly, consistently, and objectively than human vision. Computer vision was used to assess fruit quality, classify ripeness levels, detect plant diseases, and measure the physical characteristics of harvested crops [5]**.** The advantages of computer vision are that it is non-destructive, real-time, and can process large numbers of samples in a relatively short time. Application of computer vision in sorting and grading oil palm FFBs is considered highly promising because the FFB grading process remains one of the stages requiring significant human involvement. With advancements in digital cameras and modern computing techniques, such as machine learning, FFB quality parameters can be evaluated automatically with a higher level of consistency. Therefore, the development of a computer vision-based sorting and grading system is a promising alternative for improving the efficiency and productivity of the palm oil industry [6].

You Only Look Once (YOLO) is a deep learning algorithm for object detection. Machine learning and deep learning are used with computer vision systems to enable non-destructive, automatic, real-time sorting and grading of fruits and vegetables, including oil palm FFBs. Deep learning is a subset of machine learning and uses more complex algorithms. Both methods enhance the ability of computer vision systems to perform automatic image analysis. One important deep learning algorithm is a Convolutional Neural Network (CNN), which has broad applications in computer vision, especially in fresh fruit production [7]. CNNs are used for object detection, especially the Faster Region-based CNN (R-CNN). Another object detection algorithm is YOLO (You Only Look Once). In recent years, the YOLO algorithms have become one of the most popular object detection methods due to its ability to perform object detection directly in a single processing stage, resulting in high inference speeds [8]. YOLO models predict both the location and class of an object simultaneously within a single network. This approach makes YOLO more computationally efficient and suitable for real-time applications. Since its initial introduction, the YOLO algorithms have continued to evolve, resulting in various versions with improved accuracy and detection performance. These capabilities have led to YOLO being widely applied in classification and evaluating fruits and vegetables, such as for tomato ripeness identification [9].

YOLO algorithms have recently been used for oil palm FFB classification and detection. Computer vision with YOLO algorithms has been proposed for oil palm FFB classification and detection, on the tree or after harvest. An optimized YOLOv3 was used to detect and localize three objects at a palm oil plantation under various environmental conditions [10]. Another application used several YOLOv5 algorithms for FFB ripeness classification and compared them [11]. Computer vision and YOLOv4 were the first steps toward a real-time robotic FFB harvesting system [12]. Multispectral imaging and deep learning were also developed using YOLOv4 for ripeness classification [13]. YOLOv8 is one of the YOLO algorithm series that has also been used for fruit ripeness classification [14] and to detect and classify oil palm fruit bunches under real-world plantation conditions for harvesting purposes [15].

A combination of a YOLOv8-based object detection and segmentation algorithm is crucial for detecting fruit and vegetables of different sizes. The YOLOv8 object detection and instance segmentation algorithm, combined with geometric shape-fitting techniques on 3D point cloud data, enabled determination of fruit sizes, especially fruit clusters. YOLOv8s is a compromise object detection algorithm that balances accuracy, speed, and a small-sized model [16].

Another YOLOv8 family member for segmentation and classification is YOLOv8-Seg, one of the latest developments that integrated object detection and instance segmentation capabilities into a single architecture. Unlike conventional object detection models that only produce bounding boxes, YOLOv8-Seg also generates segmentation masks that show the shape of objects in greater detail. This capability is particularly important in applications involving overlapping objects or occlusion, as the system can distinguish each object individually 17]. In addition to its strong segmentation capabilities, YOLOv8-Seg also offers fast inference performance, making it suitable for real-time conveyor-based inspection systems. This model is available in various variants, ranging from YOLOv8n-Seg to YOLOv8x-Seg, each with different numbers of parameters and network complexities. Application of YOLOv8 and Mask R-CNN for orchard complex environments showed that YOLOv8 performed better than Mask R-CNN [18].

Sorting and grading oil palm FFBs at palm oil mills have some criteria. The sorting parameters include ripeness levels, minimum mass, and defects. In addition to ripeness level, the presence of physical defects in oil palm FFBs is also a key factor affecting the quality of raw materials entering the mill. These defects include empty, rotten, long stalks, excessive thorns, or spike bunches. Sorting and grading the defective FFBs are usually performed by experienced graders. Some palm oil mills have a conveyor for the sorting process, with graders lined up beside it. The defective FFB separation must be conducted at the sorting stage so that FFBs that do not meet the mill's standards do not enter the production process. However, the defect identification process is manual and traditional, which is prone to subjectivity, fatigue, and inefficiency. An automatic defective FFB identification process using computer vision and YOLOv4 has been developed; however, the FFBs are moved on the conveyor, lined up one by one in the conveyor buckets [19]. In real conditions, more than one FFB is inside the conveyor bucket or in the camera field of view (FOV); hence, identification faces various challenges. Some physical parameters that serve as indicators of defects may be obscured by other FFBs. These conditions make detection more difficult than for single objects observed separately.Conveyor movement adds complexity, as the system must detect quickly and accurately within a very short time frame. Therefore, a model capable of recognizing objects under conditions of occlusion, varying orientations, and continuous positional changes is required. The instance segmentation capability of YOLOv8-Seg is expected to help address these challenges.

In this study, normal and defective FFBs were detected using a computer system and two detection model variants YOLOv8l-Seg and YOLOv8x-Seg. Both YOLOv8l-Seg and YOLOv8x-Seg are Ultralytics YOLOv8 algorithms for instance segmentation and classification, used for large model file sizes and high accuracy. These algorithms are suitable for large-sized and stacked FFBs. Stacked FFBs mean that each conveyor bucket has one to three FFBs with the same or different defects. The performance of both models is compared. Studies

ISOMAse
International Society of Ocean, Mechanical and Aerospace Scientists and Engineers

Journal of Ocean, Mechanical and Aerospace
-Science and Engineering-
30 July 2026. Vol.70 No.2

DOI:
10.36842/jomase.v70i2.638

specifically addressing the identification of defective FFB in stacked and moving conditions on a conveyor remain very limited. In fact, these conditions are the most common scenarios encountered in the grading process within the palm oil industry. The limitations of previous research indicate that there is still a need to develop methods capable of functioning optimally in more complex environments. Most previous studies have focused on ripeness classification or object detection under relatively simple conditions and have not fully represented real-sorting operational conditions. Evaluation must not only be based on detection accuracy but also consider inference speed and object segmentation capabilities.

This study particularly aims to compare the performance of the YOLOv8-Seg and YOLOv8x-Seg models in detecting and segmenting normal and defective oil palm FFB stacked on a moving conveyor using an RGB camera-based computer vision system. The comparison analyzed various evaluation parameters such as accuracy, precision, recall, F1-score, and mean Average Precision (mAP). The findings of this study are expected to contribute to the development of more efficient, consistent, and accurate automatic grading technology in the palm oil industry. In addition, this study is expected to serve as a reference for the development of deep learning-based inspection systems for agricultural and other industrial applications.

2. MATERIALS AND METHOD

This study employs a computational experimental method by developing a YOLOv8-based computer vision system to detect the conditions of stacked fresh fruit bunches (FFB) on a moving conveyor. This system is designed to detect five classes of oil palm FFB: normal, rotten, empty bunches, long stalk, and thorny bunches. The research consisted of several stages, including the design of the computer vision systems, image acquisition, image annotation, image augmentation, YOLOv8 model construction, training, testing, and performance analysis using a confusion matrix.

2.1 Computer Vision System

The computer vision system consists of a color (RGB) camera, camera lens, camera holder, a belt conveyor unit with a motor, and a laptop. The system used daylight as a light source. The camera and lens were mounted perpendicularly above the conveyor belt at a distance of approximately 60 cm from the conveyor surface. The camera captured images and videos of oil palm fruit bunches (FFBs) during the acquisition stage. This study used a USB camera, Imaging Source DFK 33UX174 RGB, with a resolution of 1,920 × 1,200 pixels (2.3 megapixels) and a variable lens with a focal length of 4 to 12 mm. The laptop served as the primary device for connecting the camera, running image acquisition, and processing and training/testing the YOLOv8 model. Additionally, the conveyor belt transported FFB samples during video recording for model validation.

2.2 Image Acquisition and Dataset Preparation

Image acquisition was performed using an RGB camera with the JPG file format. The acquired images were then annotated using bounding boxes according to object classes. Subsequently, image augmentation, including rotation, flipping, scaling, and brightness adjustment was performed to increase dataset diversity and number. The primary research materials were oil palm fruit bunches, normal and defective bunches. The defective FFB samples used were divided into several categories: rotten, empty bunches, long stalks, and thorny bunches. This grouping was based on the physical conditions of the FFBs commonly encountered during the sorting process at oil palm mills. Figure 1 shows the visual characteristics of each sample category.

Figure 1 shows the characteristics of five oil palm bunch classes. There were 10 bunches for each class. Normal bunches have clean, dense fruits, while the long stalk bunches have a stalk length of more than 2 cm. Empty bunches are defined as more than half the fruits are loose, while rotten bunches are empty bunches that have deteriorated. The classification is based on the sorting and purchase standard for smallholders' plantation incoming FFBs at the palm oil mill. At the acquisition stage, the oil palm bunch samples were arranged in stacked conditions with variations of 2 to 3 bunches within the camera (FOV) to create conditions of occlusion and overlap between objects. This setup aimed to ensure that the acquired data accurately represents real-world field conditions, thereby enabling the developed model to achieve better detection capabilities across various bunch positions and arrangements. Figure 2 shows the arrangement of the stacked bunches to build the detection model.

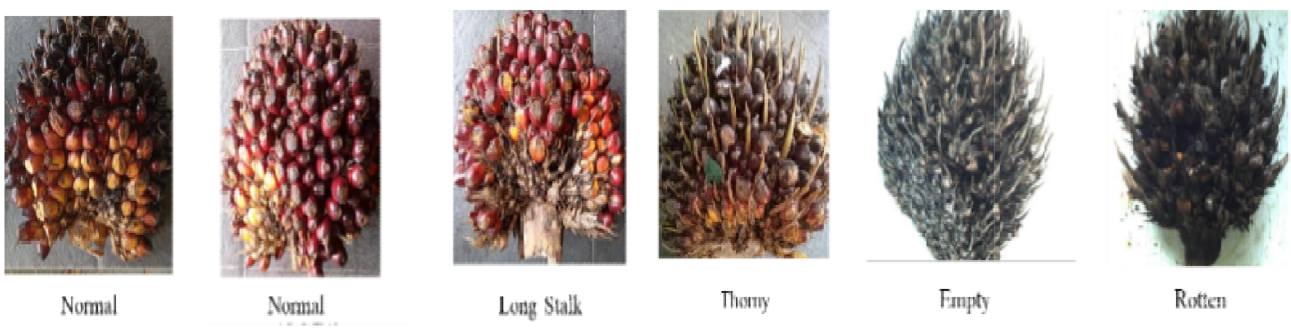
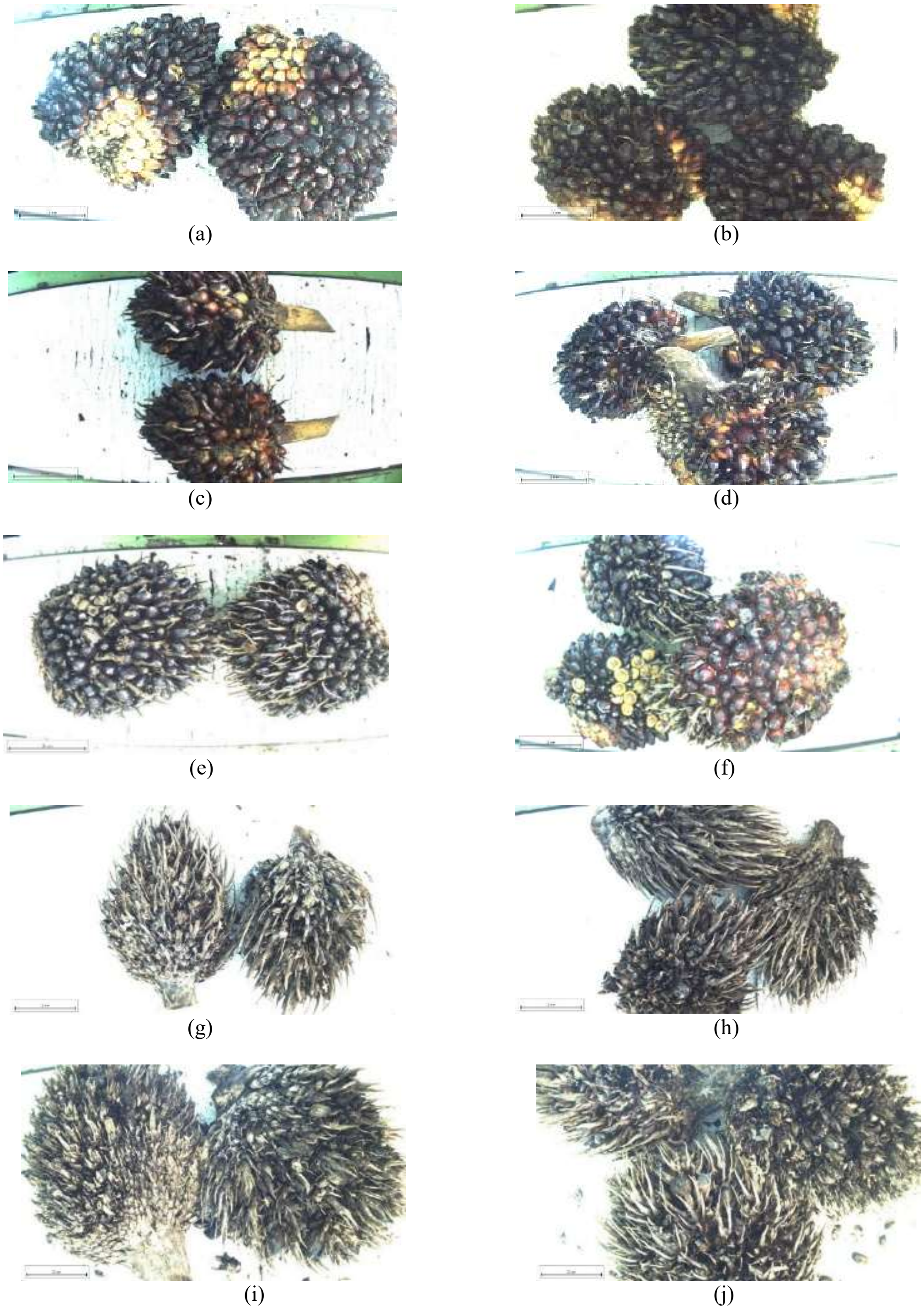


Figure 1: Sample categories of oil palm bunches

JOMAse | Received: 04-June-2026 | Accepted: 30-July-2026 | [(70) 2: 193-206]
Published by International Society of Ocean, Mechanical and Aerospace Scientists and Engineers, www.isomase.org., ISSN: 2354-7065 & e-ISSN: 2527-60



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Figure 2: Arrangement of stacked oil palm bunches

Figure 2 shows the arrangements of stacked oil palm bunches used in this study. Collection of the stacked bunches are: (a) 2 FFBs of the normal class, (b) 3 FFBs of the normal class, (c) 2 FFBs of the long-stalk class, (d) 3 FFBs of the long-stalk class, (e) 2 FFBs of the long-spike/thorny class, (f) 3 FFBs of the long-spike class, (g) 2 empty-bunch FFB, (h) 3 empty-bunch FFB, (i) 2 rotten FFB, (j) 3 rotten FFB. All images then underwent annotation to assign class labels and object location information for use in training the YOLOv8 model. The annotation process used the Roboflow application, which produced output in file.txt format. It contains the object class, image dimensions, bounding box, and polygon coordinates. After the annotation process, augmentation was also performed to increase the model dataset. In system development and data processing, this study uses the Python programming language because it offers a variety of libraries that support digital image processing and deep learning. The data acquisition process yielded a total of 545 original images used as the initial dataset.

After augmentation, the dataset size increased, and it was then divided into 80:20 for the training and testing process.

2.3 Model Construction and Validation

The detection models for identifying normal and defective oil palm FFBs in this study used the YOLOv8 algorithms, specifically YOLOv8l-seg and YOLOv8x-seg. The hyper-parameters of the YOLOv8 configuration file used include batch size (8), epochs (20), image size (640 x 640), and classes (5). Training was done previously on a computer with an HP AMD Athlon Silver 3050U, 8 GB RAM, under Windows 11 Pro 64-bit. Before the training process, a YAML data file was made, which contains the dataset and model configuration. After this stage, the training and testing were performed using the Google Colab platform. Testing was done using the remaining 20% of the dataset. During this stage, the model was evaluated and fine-tuned until reaching the highest accuracy.

The next stage in detection model construction was validation. Validation was essential after obtaining a robust model from the training and testing stages. Validation and testing are two interchangeable terms used in machine learning. Some researchers used the same dataset, which was divided into a 70:20:10 subset, 70% for training, 20% for testing, and 10% for validation.

In this study, dataset was obtained through acquisition, annotation, and augmentation processes, which were divided into 80:20 subsets. The testing dataset was used to fine-tune the hyper-parameters and to evaluate the model performance. The validation dataset was obtained separately using videos of oil palm FFBs with 5 classes moving on a conveyor belt in stacked conditions. The successfully captured images are then processed using a Python-based program with the YOLOv8-Seg model to perform automatic object detection and segmentation. Through this process, each detected FFB is assigned a bounding box, a segmentation area (mask), a class label, and a confidence score indicating the model's level of certainty regarding the prediction result.

2.4 Performance Analysis

There are at least two main parameters used for model performance analysis: average precision (AP) and mean average precision (mAP). AP and mAP values are obtained using a confusion matrix and are commonly used for deep learning models. They are derived from precision-recall curves, as given by Equations (1) and (2). Both parameters depend on the confusion matrix, which has four combinations of predicted and actual values: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). These parameters are used to calculate accuracy, precision, recall, F1-score, Average Precision (AP), and mean Average Precision (mAP).

$$AP = \sum(recall\ (n+1) - recall\ (n))\ x\ precision\ x\ recall\ (n+1) \quad (1)$$

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP(k) \quad (2)$$

3. RESULT AND DISCUSSION

This study aims to develop a real-time detection system for normal and defective oil palm fresh fruit bunches (FFBs) in stacked conditions on a conveyor belt using the YOLOv8, YOLOv8l-Seg, and YOLOv8x-Seg models. The detected defect categories include rotten, empty bunches, long stalk, and long spines or thorny bunches. The detection and segmentation processes were performed using video data as validation for the model. Image acquisition and processing were developed using the OpenCV library with the Python 3.9.5 programming language.

3.1 Annotation and Augmentation Results

The annotation process was performed manually by marking the position and category of the oil palm fruits according to the objects in the images, as shown in Figure 3. A total of 545 image files were used in this annotation stage. The result of this process produced annotation data as shown in Figure 3. The annotated sample images are then saved in YOLO annotation format as .txt files, as shown in Figure 3. Each image has one .txt file containing the bounding box coordinates and the object's polygon coordinates. Once all images have been annotated, the next step is the data augmentation process, as shown in Figure 4.

Figure 4 illustrates various data augmentation conditions for initial oil palm FFB images. Figure 4(a) shows the original image acquired using the camera. Next, Figure 4(b) shows the image rotated 90° clockwise and counterclockwise, and inverted, to improve the model's ability to recognize objects under various camera orientations. Furthermore, Figure 4(c) displays the results of rotations of -5° and +5°, aimed at enabling the model to detect objects even when the camera or object positions are not perfectly aligned. In Figure 4(d), brightness adjustments were applied to darken the image by -13% and brighten it by +13% to help the model adapt to variations in lighting and camera settings. All of these augmentation results were then used during the model training phase. The total number of annotated and augmented images is summarized in Table 1.

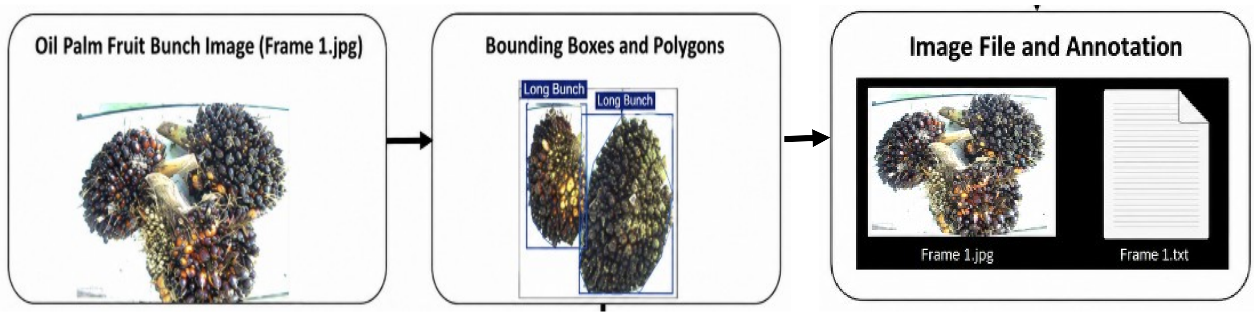


Figure 3: Results of the dataset annotation process using Roboflow

Journal of Ocean, Mechanical and Aerospace
-Science and Engineering-
30 July 2026. Vol.70 No.2
DOI:
10.36842/jomase.v70i2.638

(a)

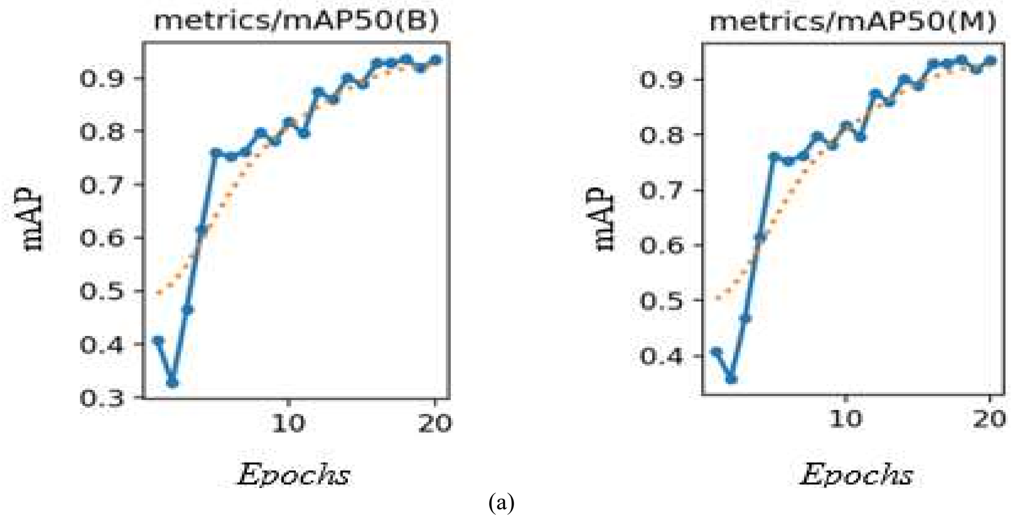
(b)

(c)

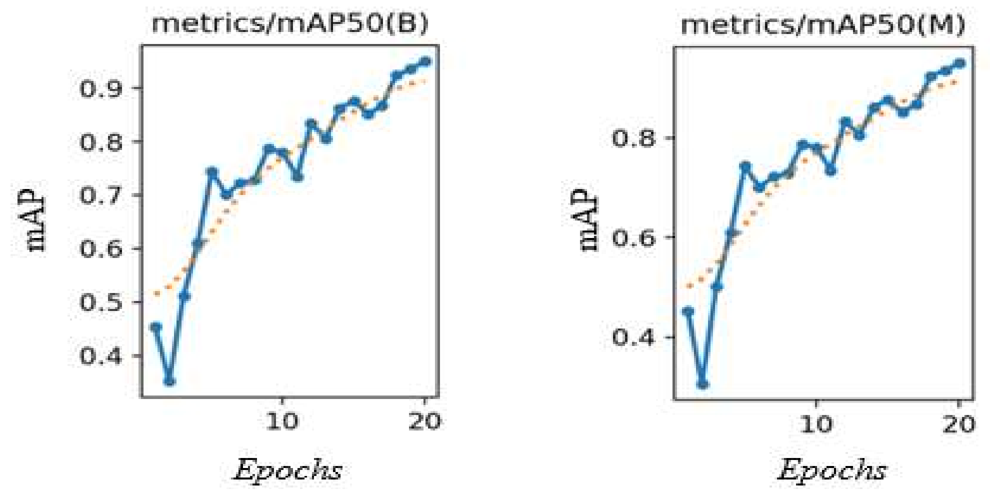
(d)

Figure 4: Results of the augmentation process: (a) Original image, (b) Image rotated 90°, (c) Rotation, (d) Brightness

ISOMAse
International Society of Ocean, Mechanical and Aerospace Scientists and Engineers
Journal of Ocean, Mechanical and Aerospace
-Science and Engineering-
30 July 2026. Vol.70 No.2
DOI:
10.36842/jomase.v70i2.638



(a)



(b)

Figure 5: Training results for 20 epochs: (a) YOLOv8l-Seg model; (b) YOLOv8x-Seg model

The total number of images in Table 1 was obtained by manually counting all oil palm FFB objects present in each frame. This process was performed to prevent data imbalance and reduce detection errors in the model. Each category, namely, normal, long stem, many thorns, empty bunch, and rotten that was contained 261 images. Subsequently, the dataset was split into a subset of 80% for training and 20% for testing.

Table 1: Total number of images after augmentation

Class	Number of Images
Normal	261
Long spike/thorny	261
Long Stalk	261
Empty Bunch	261
Rotten	261
Total	1305

3.2 Training Results of YOLOv8 Model

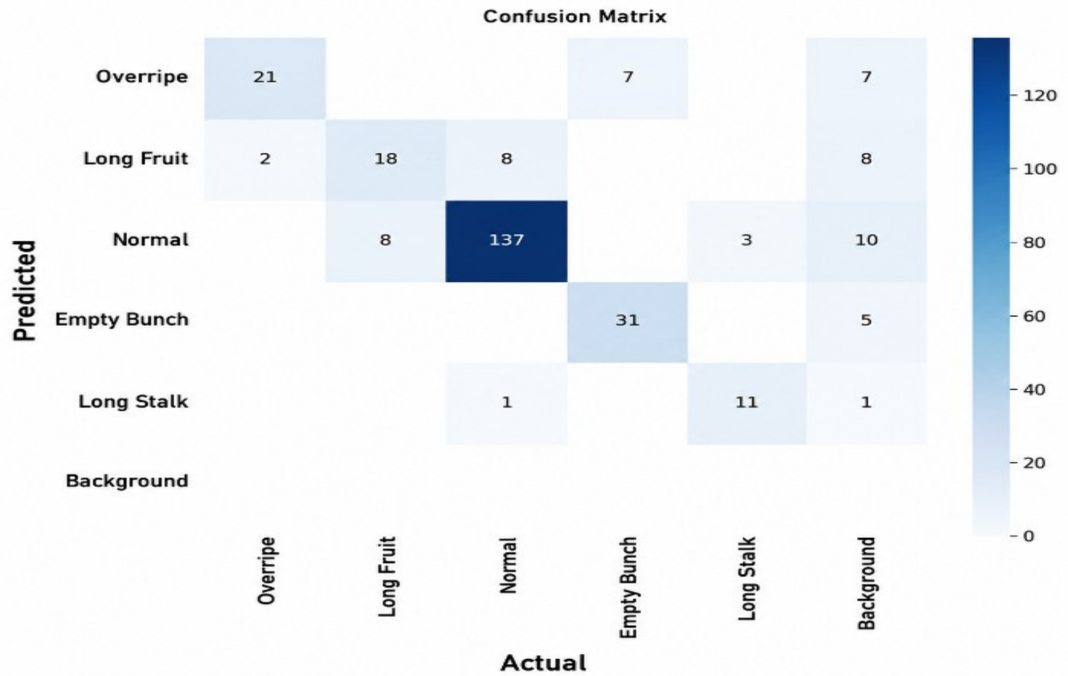
The model training process was conducted over 20 epochs with a batch size of 8 and an input image size of 640×640 pixels. This training step aims to train the constructed model to recognize object patterns based on annotated images, enabling the model to generate more accurate detection decisions. Then, data augmentation was performed using Roboflow to generate image variations, thereby improving the model's generalization ability. The training process was conducted using the Google Colab cloud-based platform, which is easier to access via a browser without requiring additional installation. In this study, two YOLO models were used: YOLOv8 (specifically the YOLOv8l-seg variant) and YOLOv8x-seg, resulting in training performance graphs as shown in Figure 5.

Figure 5 shows the results of the YOLOv8 model training process, with the x-axis representing number of epochs and the y-axis showing the mAP (mean Average Precision) value. The observed parameters include mAP50(B) and mAP50(M). The mAP50(B) value describes the model's ability to detect objects using bounding boxes at an IOU of 0.5. At the same time,

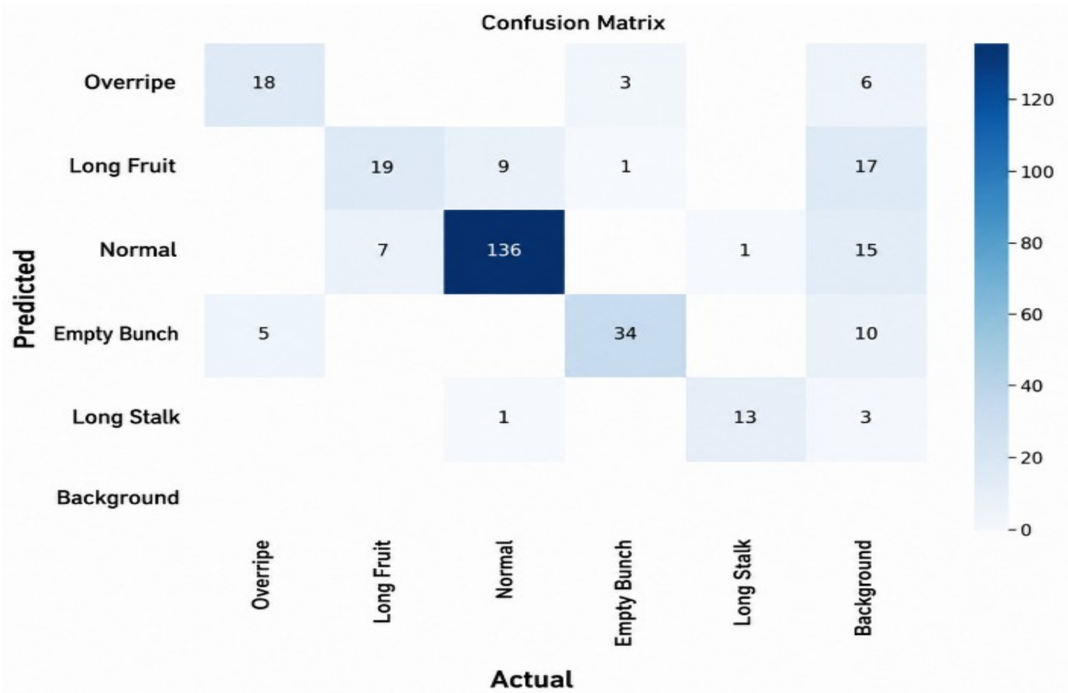
ISOMAse
International Society of Ocean, Mechanical and Aerospace Scientists and Engineers
Journal of Ocean, Mechanical and Aerospace
-Science and Engineering-
30 July 2026. Vol.70 No.2
DOI:
10.36842/jomase.v70i2.638

mAP50(M) indicates the performance of object segmentation using masks at the same IOU value. The graphs in Figures 5(a) and 5(b) show a comparison of the performance of the YOLOv8l-Seg and YOLOv8x-Seg models over 20 epochs.

Both models show rapid performance improvement early in training, with the mAP value increasing from around 0.35 to over 0.90 by the end of training. However, YOLOv8l-Seg exhibits sharper fluctuations over several epochs, whereas YOLOv8x-Seg shows a more stable and consistent upward trend. The orange dashed line in the graph indicates the baseline performance value used as a reference during the training process. For the mAP50(M) metric, YOLOv8l-Seg experienced noticeable fluctuations between epochs 10 and20, indicating lower training stability.



(a)



(b)

Figure 6: Table of confusion matrix analysis results using test data: (a) YOLOv8l-Seg, (b) YOLOv8x-Seg

Table 2: Results of the YOLOv8 model training

Model YOLO	mAP50(B)%	mAP50(M)%
YOLOv8l-Seg	90	90
YOLOv8x-Seg	90	80

Table1: Performance analysis using confusion matrix

Model	Accuracy %	Precision %	Recall %	F1-score %
YOLOv8l-Seg	92%	76%	86%	78%
YOLOv8x-Seg	93%	70%	86%	76%

In contrast, YOLOv8x-Seg demonstrated a more stable improvement in segmentation performance, although the increase was more gradual. Both models achieved an mAP50(M) value of around 50% by the end of training. The fluctuating performance of YOLOv8l-Seg suggests a potential for overfitting or underfitting, whereas YOLOv8x-Seg is considered better at adjusting parameters and learning data patterns during the training process. The results are summarized in Table 2.

3.4 Performance Analysis of YOLO Models

The confusion matrix is one of the evaluation methods used to measure the performance of a detection model. There are four main parameters in the confusion matrix: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). These four parameters form the basis for calculating various evaluation metrics, such as accuracy, precision, recall, and F1-score, to assess the model's ability to distinguish each class accurately. The evaluation results are shown in Figure 6.

In Figure 6 shows the results of calculating the number of stacked oil palm fresh fruit bunches (FFB) in an image using a confusion matrix. Based on these results, the YOLOv8l-Seg model produced a higher False Positive (FP) rate compared to YOLOv8x-Seg. Meanwhile, YOLOv8x-Seg showed fewer false positives, indicating better performance in minimizing prediction errors. This difference indicates that the larger YOLO model tends to be more sensitive to objects, yet still delivers more optimal detection performance. Detection errors in the form of false positives can be influenced by several factors, such as suboptimal lighting conditions, low image quality, and non-uniform object colors [20]. Based on the results of the confusion matrix, the accuracy, precision, recall, and F1-score can then be calculated, as shown in Table 3.

Accuracy, precision, recall, and F1-score are metrics used to evaluate the performance of a detection model. Accuracy indicates the model's ability to make correct predictions across the entire test dataset. Precision describes the model's ability to generate correct positive predictions, while recall indicates the model's ability to detect all positive data. Meanwhile, the F1-score combines precision and recall to assess the model's performance balance. Based on Table 3, the YOLOv8l-Seg model achieved an accuracy of 92% in detecting normal and defective oil palm FFB, while YOLOv8x-Seg achieved a slightly higher accuracy of 93%. The precision values obtained by YOLOv8l-Seg and YOLOv8x-Seg were 76% and 70%, respectively. In terms of processing speed, YOLOv8l-Seg is capable of performing detection in 0.5 ms, while YOLOv8x-Seg requires approximately 3.9 ms. This indicates that the larger the model size used, the longer the computation time required.

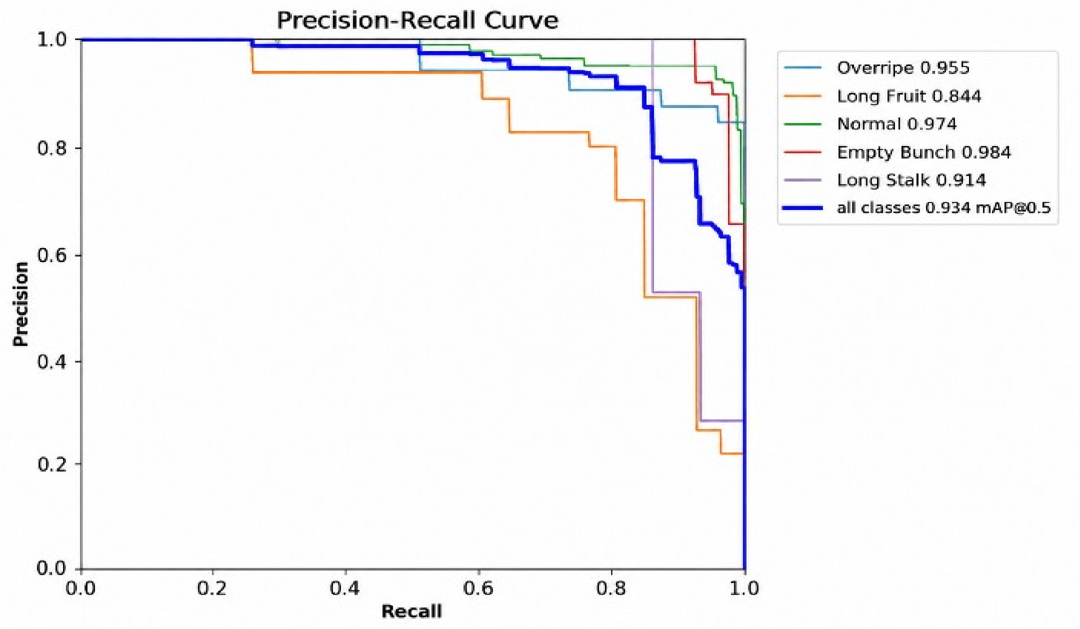
The evaluation metrics used to assess the performance of a machine learning algorithm must be tailored to the dataset's characteristics and detection objectives. Accuracy is appropriate when the number of false positives and false negatives is relatively balanced. However, if the two are unbalanced, the F1-score is a more suitable evaluation metric. Recall is prioritized when false negative errors are minimized. At the same time, precision is used when the system is expected to produce a high number of true positives and reduce false positives. Additionally, high specificity is required to minimize false positive errors [21].

A study was conducted on detecting oil palm fresh fruit bunches (FFBs) piled on the ground using the YOLOv4-q-fp16 model for the categories of empty, ripe, unripe, overripe, and abnormal bunches, achieving anmAP of 65.53% [22]. Compared to that study, the YOLOv8-Seg model in this study demonstrated higher accuracy performance, particularly in the YOLOv8x-Seg model, although the improvement was not very significant compared to YOLOv8l-Seg. Based on Table 2, the mAP values for detection and segmentation indicate that YOLOv8l-Seg has more stable and consistent performance compared to YOLOv8x-Seg. Additionally, mAP calculations across various IOU thresholds were analyzed using the relationship between precision and recall curves. These curves provide a more detailed overview of the model's performance at each tested threshold. Figure 7 depicted the precision-Recall curves and Table 4 shows the YOLOv8 variant specifications.

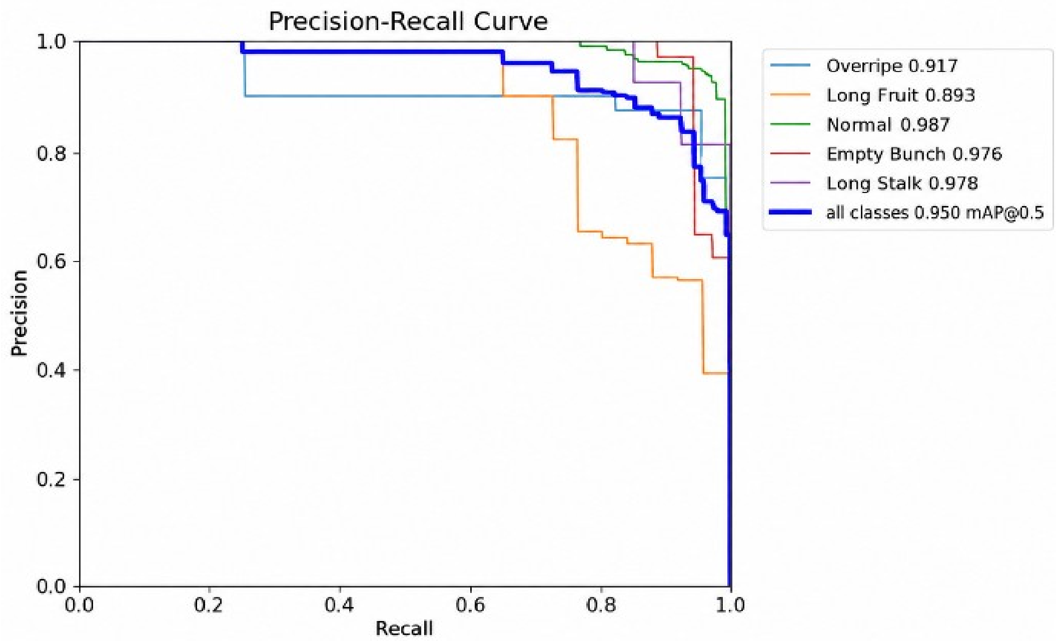
Table 4: YOLOv8 variant specifications

Model	Size (pixel)	mAP^{box} 50-95	mAP^{mask} 50-95
YOLOv8n-Seg	640	36.7	30.5
YOLOv8s-Seg	640	44.6	36.8
YOLOv8m-Seg	640	49.9	40.8
YOLOv8l-Seg	640	52.3	42.6
YOLOv8x-Seg	640	53.4	43.4

ISOMAse International Society of Ocean, Mechanical and Aerospace Scientists and Engineers
Journal of Ocean, Mechanical and Aerospace
-Science and Engineering-
30 July 2026. Vol.70 No.2
DOI: 10.36842/jomase.v70i2.638



(a)



(b)

Figure 7: Precision-recall curves: (a) YOLOv8l-Seg, (b) YOLOv8x-Seg

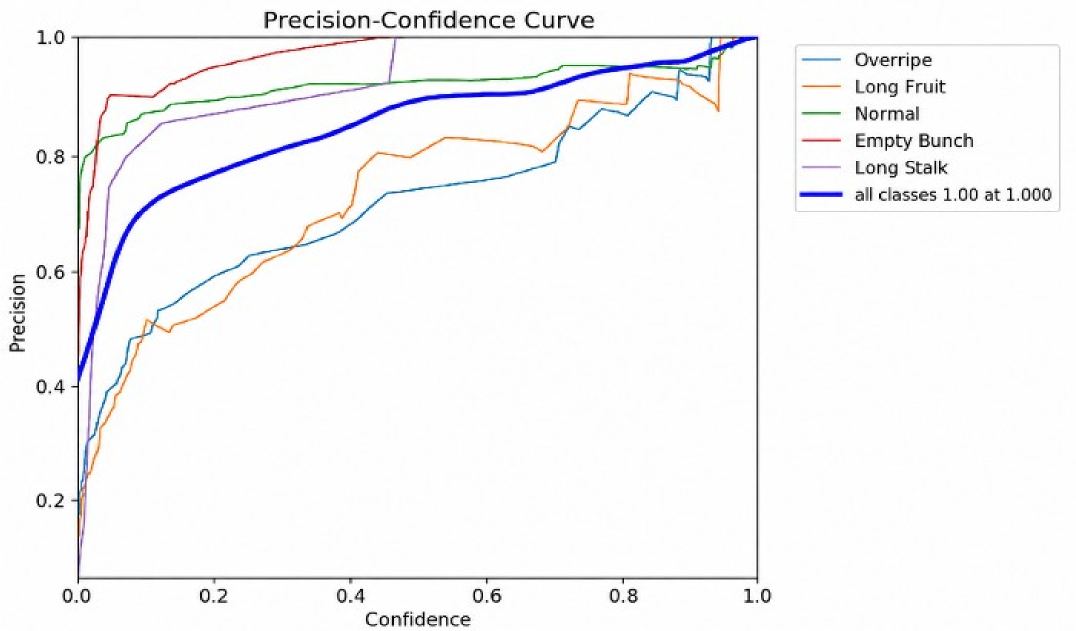
Figure 7 shows the relationship between precision and recall for the two models used. Based on the graph, the YOLOv8l-Seg model achieved anmAP of 93.4%, which is the average AP (Average Precision) across all sample classes. Meanwhile, YOLOv8x-Seg produces a higher value, namely 95%. Figure 7 also shows that the green line representing the normal class has a higher value compared to the long stem, many thorns, empty bunch, and rotten classes. This indicates that both models have excellent capabilities in detecting stacked oil palm FFB in the normal category. Table 4 shows the standard to evaluate the resulting detection models.

Based on COCO (Common Objects in Context) data, which serves as the evaluation standard for detection models, Table 4 shows that the YOLOv8x-Seg model achieved anmAP score of 53.4 [23]. The results show that YOLOv8x-Seg performs better than YOLOv8l-Seg in detecting objects in images. The precision values for each sample class in Figure 7 show that YOLOv8l-Seg achieved 97% precision for the normal class, 91% for long stalks, 84% for thorny bunches, 98% for empty bunches, and 95% for the rotten class. Meanwhile, YOLOv8x-

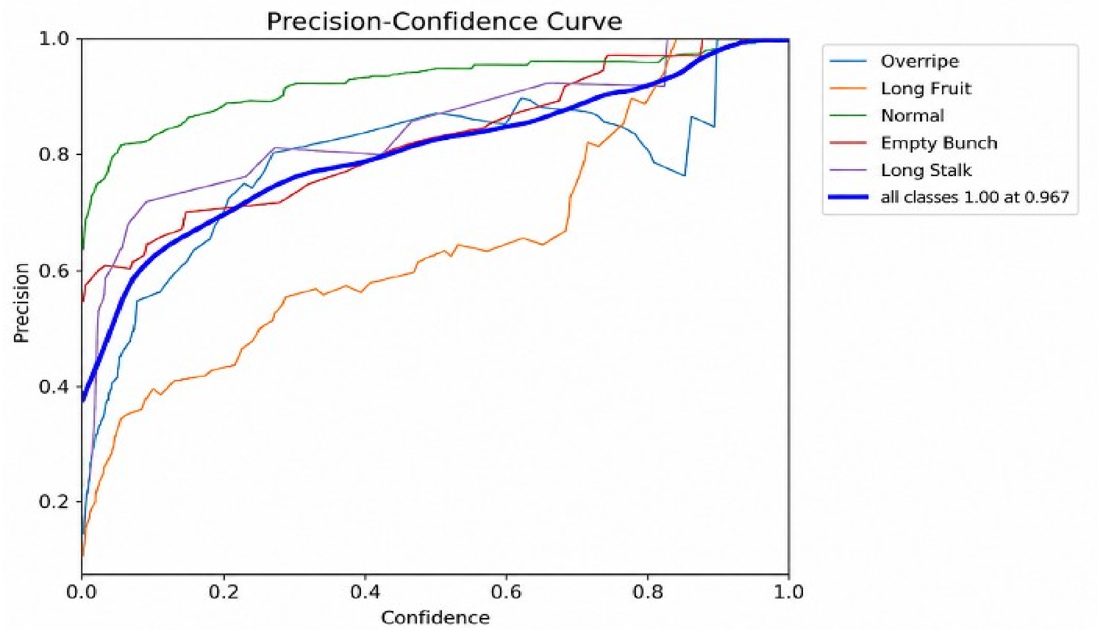
Journal of Ocean, Mechanical and Aerospace
-Science and Engineering-
30 July 2026. Vol.70 No.2
DOI: 10.36842/jomase.v70i2.638

Seg achieves a precision of 98% for the normal class, 97% for long stems, 89% for long spines, 97% for empty bunches, and 91% for the rotten class. These high-precision values indicate that both models can distinguish palm oil FFB classes—such as normal, long stems, long spines, empty bunches, and rotten that was quite well. Figure 8 illustrates the relationship between precision and confidence in the detection model.

Figure 8 shows the relationship between precision and confidence in the detection model. Precision measures the model's accuracy in labeling objects, while confidence indicates the model's level of certainty about the predicted results. In this study, the YOLOv8l-Seg and YOLOv8x-Seg models achieved a precision score of 100%, indicating that the models are capable of labeling objects with high accuracy. This states that an increase in the precision score generally accompanies an increase in the confidence score. Models with high precision scores demonstrate a strong ability to minimize false positives. However, the model may still fail to detect all positive objects if the recall score is not yet optimal.



(a)



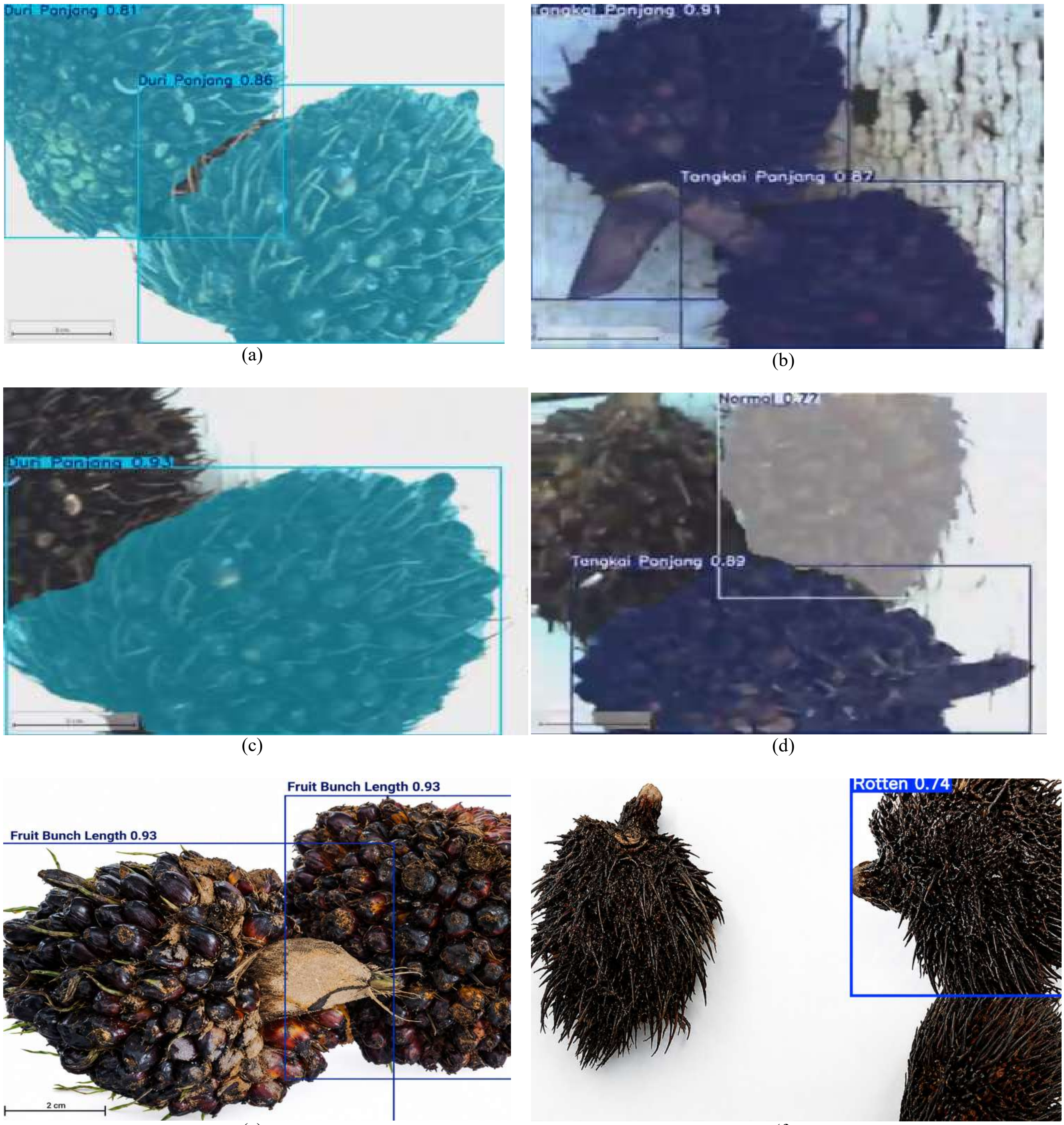
(b)

Figure 8: Precision-confidence curves: (a) YOLOv8l-Seg, (b) YOLOv8x-Seg

ISOMAse
International Society of Ocean, Mechanical and Aerospace Scientists and Engineers
Journal of Ocean, Mechanical and Aerospace
-Science and Engineering-
30 July 2026. Vol.70 No.2
DOI: 10.36842/jomase.v70i2.638

3.5 Validation Result Using Frame Video

After the training stage, the model was tested using a Python- and OpenCV-based detection system with 20% of the images. Model validation used video frames. This stage is known as the prediction process, which was conducted using video footage recorded by an RGB camera. During this process, the YOLO model's performance was evaluated. The evaluation compared the model's predictions against the ground-truth labels in the test data. Evaluation metrics such as precision, recall, and F1-score were used to measure the model's effectiveness in distinguishing between normal and defective FFBs. If an object is successfully detected, the system displays a bounding box and segmentation of the object along with a confidence score and class label, such as normal, long stalk, thorny, empty bunch, or rotten. Figure 9 shows the validation results.



(a) (b) (c) (d) (e) (f)

Figure 9: Testing the detection of oil palm fruit ripeness using video.

JOMAse | Received: 04-June-2026 | Accepted: 30-July-2026 | [(70) 2: 193-206]
Published by International Society of Ocean, Mechanical and Aerospace Scientists and Engineers, www.isomase.org., ISSN: 2354-7065 & e-ISSN: 2527-60

Figure 9 shows the results of video frames tested using the YOLOv8 and YOLOv8x-Seg models. In this test, oil palm fresh fruit bunches (FFBs) were placed on a moving conveyor belt in a mixture of normal and defective conditions, with varying positions. Figure 9(a) shows that the FFB was successfully and accurately detected as the "thorny" class, while Figure 9(b) demonstrates accurate detection in the "long stalk" class. In Figure 9(c), the object was successfully detected, but the segmentation did not cover the entire FFB. Figure 9(d) shows that two FFB were correctly detected, while one other object was not detected. In Figure 9(e), one FFB was correctly identified, but there was an error in labelling the other object. Furthermore, Figure 3.9(f) shows that only one FFB was correctly detected, while several other objects were not detected.

The validation process was conducted using a Python script that input a video file; the system then displayed the bounding box, segmentation, class labels, and confidence scores for each detected object. Each class was assigned a different border color, such as white for normal FFB, dark blue for long stalks, light blue for long spines, and dark blue for rotten FFB. Testing was conducted on FFB of various sizes and orientations, resulting in varying confidence scores. Based on the test results in Figure 9, the system was able to detect 1 to 3 oil palm FFBs fairly well, complete with bounding boxes and object segmentation. However, under certain conditions, the system still experiences errors in class labeling as well as detection failures. This is influenced by the FFBs being too close together and the similarity in visual characteristics between normal and defective FFBs in the training dataset, causing the model to struggle to distinguish objects accurately.

4. CONCLUSION

This study developed an effective detection method for identifying normal and defective oil palm fresh fruit bunches (FFB) using the YOLOv8-Seg model. The YOLOv8-Seg model demonstrated superior accuracy, achieving 93%, with a processing time of approximately 3.9 ms (milliseconds). Meanwhile, YOLOv8-Seg achieved an accuracy of 92% with a faster detection time of 0.5 ms. Although both models performed quite well, the system still had difficulty distinguishing between certain FFB classes, such as normal, long-stalk, long-spike, empty bunch, and rotten that was due to the similar color characteristics among the samples. Future studies should focus on improving YOLOv8-based detection by increasing dataset diversity and incorporating advanced image-processing or deep-learning techniques to better distinguish FFB classes with similar visual characteristics. Further validation under varying conveyor speeds, lighting conditions, stacking configurations, and real-time industrial environments is also recommended to support reliable automated FFB sorting.

REFERENCES

[1] Lai, J. W., Ramli, H. R., Ismail, L. I. & Wan Hasan, W. Z. 2023). Oil Palm Fresh Fruit Bunch Ripeness Detection Methods: A Systematic Review. Agriculture (Switzerland), 13(1). https://doi.org/10.3390/agriculture13010156.

[2] Supriyatin, W. (2021). Palm oil extraction rate prediction based on the fruit ripeness levels using C4.5 algorithm. *ILKOM Jurnal Ilmiah*, *13*(2), 92–100. https://doi.org/10.33096/ilkom.v13i2.714.92-100.

[3] Mansour, M. Y. M. A., Dambuli, K. D. & Yeep1, C. K. (2022). A review of non-destructive ripeness classification techniques for oil palm fresh fruit bunches. *Journal of Oil Palm Research*, 35(4), 543–554. https://doi.org/10.21894/jopr.2022.0063.

[4] Tian, H., Wang, T., Liu, Y., Qiao, X. & Li, Y. (2020). Computer vision technology in agricultural automation —A review. In *Information Processing in Agriculture* (Vol. 7, Nomor 1). https://doi.org/10.1016/j.inpa.2019.09.006.

[5] Ireri, D., Belal, E., Okinda, C., Makange, N. & Ji, C. (2019). A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. *Artificial Intelligence in Agriculture*. https://doi.org/10.1016/j.aiia.2019.06.001.

[6] Ismail, N. & Malik, O. A. (2022). Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. *Information Processing in Agriculture*, *9*(1), 24–37. https://doi.org/10.1016/j.inpa.2021.01.005.

[7] Wang, C., Liu, S., Wang, Y., Xiong, J., Zhang, Z., Zhao, B., Luo, L., Lin, G. & He, P. (2022). Application of convolutional neural network-based detection methods in fresh fruit production: a comprehensive review. In *Frontiers in Plant Science* (Vol. 13). https://doi.org/10.3389/fpls.2022.868745.

[8] Gao, F., Yang, T. & Fu, L. (2021). Apple fruit detection and counting based on deep learning and trunk tracking. *American Society of Agricultural and Biological Engineers Annual International Meeting,* ASABE 2021, 1. https://doi.org/10.13031/aim.202100193.

[9] Xiao, B., Nguyen, M. & Yan, W. Q. (2023). Fruit ripeness identification using YOLOv8 model. *Multimedia Tools and Applications*, 83(9), 28039–28056. https://doi.org/10.1007/s11042-023-16570-9.

[10] Junos, M. H., Mohd Khairuddin, A. S., Thannirmalai, S. & Dahari, M. (2021). An optimized YOLO-based object detection model for crop harvesting system. *IET Image Processing,* 15(9), 2112–2125. https://doi.org/10.1049/ipr2.12181.

[11] Mansour, M. Y. M. A., D. Dambul, K. & Choo, K. Y. (2022). Object detection algorithms for ripeness classification of oil palm fresh fruit bunch. *International Journal of Technology,* 13(6), 1326. https://doi.org/10.14716/ijtech.v13i6.5932.

[12] Lai, J. W., Ramli, H. R., Ismail, L. I. & Hasan, W. Z. W. (2022). Real-time detection of ripe oil palm fresh fruit bunch based on YOLOv4. *IEEE Access*, *10*, 95763–95770. https://doi.org/10.1109/ACCESS.2022.3204762

[13] Shiddiq, M., Saktioto, Salambue, R., Wardana, F., Dasta, V. V., Harmailil, I. O., Rabin, M. F., Arpyanti, N. & Wahyudi, D. (2024). Multispectral imaging and deep learning for oil palm fruit bunch ripeness detection. *Bulletin of Electrical Engineering and Informatics*, 13(6).https://doi.org/10.11591/eei.v13i6.8120

[14] Kiet, V. M., Chang, F. S. & Tsung, C. K. (2025). A study of oil palm fruit ripeness using YOLOv8 for smart agriculture applications. IS3C 2025 - *International Symposium on Computer, Consumer and Control.*

https://doi.org/10.1109/IS3C65361.2025.11131089.

[15] Gunawan, T. S., Kartiwi, M., Mansor, H. & Yusoff, N. Md. (2023). Palm fruit ripeness detection and classification using various YOLOv8 models. *IEEE 9th International Conference on Smart Instrumentation, Measurement and Applications* (ICSIMA), 193–198. https://doi.org/10.1109/ICSIMA59853.2023.10373435.

[16] Sapkota, R., Ahmed, D., Churuvija, M. & Karkee, M. (2024). immature green apple detection and sizing in commercial orchards using YOLOv8 and shape fitting techniques. *IEEE Access,* 12, 43436–43452. https://doi.org/10.1109/ACCESS.2024.3378261.

[17] Li, P., Zheng, J., Li, P., Long, H., Li, M. & Gao, L. (2023). Tomato maturity detection and counting model based on MHSA-YOLOv8. *Sensors,* 23(15), 6701. https://doi.org/10.3390/s23156701.

[18] Sapkota, R., Ahmed, D. & Karkee, M. (2024). Artificial intelligence in agriculture comparing YOLOv8 and mask R-CNN for instance segmentation in complex orchard environments. *Artificial Intelligence in Agriculture*, 13, 84–99. https://doi.org/10.1016/j.aiia.2024.07.001.

[19] Arief, D. S, Shiddiq, M., Jahrizal, J., Dalil, M., Dalil, Salambue, R., Yuhendra, A., Dasta, V. V., Hamid, M.I., Anjarwati, D. & Harmailil, M. I. O. (2026). Machine vision for detection of defective oil palm fresh fruit bunches using YOLO Algorithm, *J. Phys.: Conf. Ser.* 3186 012026.

[20] Kainec, K. A., Caccavaro, J., Barnes, M., Hoff, C., Berlin, A. & Spencer, R. M. C. (2024). Compared to research-grade actigraphy and polysomnography. *Sensors,* 24, 635.

[21] Chicco, D. & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. 1–13.

[22] Suharjito, Asrol, M., Utama, D. N., Junior, F. A. & Marimin. (2023). Real-time oil palm fruit grading system using smartphone and modified YOLOv4. *IEEE Access*, 11, 59758–59773. https://doi.org/10.1109/ACCESS.2023.3285537.

[23] Zou, Z., Chen, K., Shi, Z., Guo, Y. & Ye, J. (2023). Object Detection in 20 Years: A Survey. *Proceedings of the IEEE*, *111*(3), 257–276. https://doi.org/10.1109/JPROC.2023.3238524.