Opsi
Vol. 19, No. 1, June 2026
p-ISSN 1693-2102
e-ISSN 2686-2352

Implementation of YOLO11 for accurate waste detection

Evrita Lusiana Utari [1], Restiana Anggraeni [2], Dhimas Arief Dharmawan [2], Latifah Listyalina [3]*, Herison Surbakti [4]

[1]Department of Electrical Engineering, Universitas Respati Yogyakarta, Sleman 55281, Indonesia
[2]Department of Informatics, Universitas Pembangunan Nasional "Veteran" Yogyakarta, Sleman 55281, Indonesia
[3]Department of Rubber and Plastic Processing, Politeknik ATK Yogyakarta, Bantul 55188, Indonesia
[4]Information and Communication Technology, International College, Rangsit University, Pathum Thani 12000, Thailand

***Corresponding Author**: latifah.listyalina@atk.ac.id

Article history:
Received: 16 September 2025
Revised: 11 June 2026
Accepted: 28 July 2026
Published: 30 July 2026

Keywords:
Deep learning
Object detection
Waste classification
YOLO11n

ABSTRACT

Waste management, particularly the waste sorting process, remains a major challenge in urban areas. Manual sorting is often inefficient and prone to human error, highlighting the need for an accurate and automated waste classification system. This study investigates the implementation of the YOLO11 model for detecting and classifying six waste categories—plastic, paper, glass, metal, electronics, and organic—using a bounding box-based object detection approach. The lightweight YOLO11n variant developed by Ultralytics was employed in this study. The dataset was obtained from previous research and subsequently annotated, augmented, and divided into training, validation, and testing subsets. The model was trained for 100 epochs using the default configuration on an NVIDIA GeForce RTX 4060 Ti GPU (16 GB VRAM). During the final evaluation on the unseen test dataset, the model demonstrated reliable generalization and classification performance, achieving an overall macro-average precision of 0.75, an overall recall of 0.87, and a global F1-score of 0.79. The Paper and Plastic classes showed the highest detection performance, with Plastic achieving a flawless recall of 1.00, while the Organic class remained the most challenging due to higher visual variability. These results demonstrate that YOLO11n provides an effective approach for automated waste detection and has the potential to support more efficient and accurate waste management systems.

DOI:
https://doi.org/10.31315/opsi.v19i1.15491

1. INTRODUCTION

Waste has become an increasingly complex environmental issue [1], particularly in urban areas. According to the latest data from the Indonesia's National Waste Management Information System (SIPSN) in 2024, the daily waste generation in Yogyakarta City reached 300.56 tons [2]. The high volume of waste—

especially inorganic types such as plastic and metal—poses a serious threat to the environment as it can take hundreds of years to decompose naturally [3]. For example, microplastic pollution in aquatic systems is a progressively prevailing international problem that contributes to severe impacts on ecosystem functioning and human health [4]–[6]. Therefore, waste sorting based on type is a crucial step in sustainable waste management [7].

Efficient waste sorting is essential for improving recycling processes and reducing environmental pollution. However, in many cases, waste sorting is still carried out manually, which is often inefficient, time-consuming, and prone to human error [8]. These limitations have encouraged the adoption of advanced technologies, particularly Artificial Intelligence (AI), to support automated waste classification. The application of AI, especially machine learning (ML), has been shown to improve the accuracy and efficiency of waste classification systems [9].

Traditional machine learning methods commonly rely on sliding-window techniques to extract handcrafted visual features from images, such as SIFT [10], [11] and HOG [12]. After feature extraction, classification algorithms, including Decision Trees [13] and Support Vector Machines [14], are applied to determine whether the analyzed window contains the target object. However, the effectiveness of these approaches is often limited because they depend heavily on manually designed features.

With the advancement of deep learning, target detection methods based on neural networks have gained significant attention due to their ability to automatically learn complex visual representations without relying on handcrafted features. These models offer a simpler framework while enabling the extraction of high-level semantic information, which improves adaptability and generalization. As a result, deep learning-based object detection models have continued to evolve and improve over time. Recently, deep learning techniques have also been applied to detect and identify waste in real-world environments using drones or vehicle-mounted cameras [15]. Early deep learning detection frameworks include two-stage models such as R-CNN [16] and Faster R-CNN [17], which generally achieve higher detection accuracy but require greater computational resources.

Recently, Rahmatulloh et al. [18] implemented the YOLOv7-tiny model for waste object detection and obtained an mAP@0.5 of 86.8%. However, several limitations remain. The use of the lightweight YOLOv7-tiny architecture may also restrict the model's capability in extracting more complex visual features compared with more recent YOLO architectures. Furthermore, the experimental results showed inconsistent performance across classes, such as the glass category which achieved only 68% accuracy. These limitations indicate that further research is needed to explore more advanced object detection models for improving waste detection performance.

Recent developments in object detection models have introduced more advanced architectures such as YOLOv9, YOLOv10, and YOLO11. These YOLO variants have been evaluated on various object detection tasks using small-scale datasets [19]–[21]. More specifically, according to Jiang & Zhong [22], YOLO11 demonstrated the best detection performance on ODVerse33, a benchmark comprising 33 datasets from various domains. The model achieved higher mean Average Precision (mAP) and showed improved robustness in handling complex environmental variations compared with earlier YOLO versions, including YOLOv7-tiny. YOLO (You Only Look Once) is a widely used real-time object detection algorithm known for its high detection speed and accuracy [23]–[25]. The latest version, YOLO11, developed by Ultralytics, introduces architectural improvements that enhance detection performance while maintaining computational efficiency.

Therefore, this study proposes the use of the YOLO11 model for automated waste detection and classification using a bounding box-based object detection approach. The proposed system focuses on identifying six waste categories: plastic, paper, glass, metal, electronics, and organic waste. To balance detection performance and computational efficiency, this study employs YOLO11n, a lightweight variant of the YOLO11 architecture designed to provide efficient inference while maintaining competitive detection accuracy. Such a configuration is particularly suitable for resource-constrained environments and potential real-time waste management applications. By leveraging the improved detection capability of YOLO11, this research aims to develop a more accurate and robust waste classification system capable of identifying different waste categories in images. The main contribution of this study lies in the implementation and evaluation of the latest YOLO11 architecture for multi-class waste detection, demonstrating improved detection performance compared with earlier YOLO-based approaches.

2. METHODOLOGY

This study develops an automated waste detection system using the YOLO11n object detection model, a lightweight variant of the YOLO11 architecture that provides a balance between detection accuracy and computational efficiency. The proposed system is designed to identify six categories of waste objects in images using a bounding box-based object detection approach: plastic, paper, glass, metal, electronics, and organic waste. The use of the YOLO11n model enables efficient training and inference while maintaining reliable detection performance, making it suitable for potential real-time or resource-constrained waste management applications. The overall research workflow consists of dataset preparation, data preprocessing, model training, and performance evaluation.

2.1. Data Collection

The dataset used in this study is secondary data obtained from the work of Rahmatulloh et al. [18]. The dataset consists of six waste categories: plastic, paper, glass, metal, electronics, and organic. The first four categories (plastic, paper, glass, and metal) were collected through direct photography using mobile devices, including iPhone 7 Plus, iPhone 5S, and iPhone SE. The images were captured against a white background under both natural lighting conditions and indoor illumination, and subsequently resized to a resolution of 512 × 384 pixels.

The remaining two categories, electronics and organic waste, were collected from publicly available internet sources and also standardized with a white background to maintain visual consistency across the dataset. The dataset is publicly accessible at https://s.id/WasteIn and contains a total of 1,209 images. Each waste category contains approximately 200 images, except for the paper category which includes 199 images. In addition, 10 images without annotated objects were included as background samples. All six waste categories—plastic, paper, glass, metal, electronics, and organic—were used in this study. The distribution of images for each class is summarized in Table 1.

Table 1. Distribution of Images across waste categories *in the wastein dataset*

Class	Number of images
Electronic	200
Glass	200
Metal	200
Organic	200
Paper	199
Plastic	200
Background	10

2.2. Data Preprocessing

Before training the YOLO11n model, the dataset underwent several preprocessing steps to ensure compatibility with the object detection framework. The preprocessing stage consisted of four main processes: data annotation, data splitting, image resizing, and data augmentation. All preprocessing steps were performed using the Roboflow platform. In the annotation stage, each waste object in the dataset was labeled using a bounding box annotation format, which is commonly used in object detection tasks. This annotation defines the location and class label of the object within the image, enabling the model to learn both object classification and localization. An example of the bounding box annotation used in this study is shown in Figure 1.

In the second stage, the dataset was divided into three subsets through a data splitting process to support model training and evaluation. The dataset was partitioned into training, validation, and testing subsets using a ratio of 70%, 20%, and 10%, respectively. Allocating a larger portion of the dataset to the training subset allows the model to learn representative visual features from the available data. Meanwhile, the validation subset is used to monitor the training process and assist in tuning model parameters, while the testing subset is reserved for the final evaluation to measure the model's generalization capability on unseen data.

Figure 1. Bounding box annotation

After the data splitting process, image resizing was performed to ensure that all images met the input size requirements of the YOLO11n model. Image resizing is a preprocessing step that adjusts the spatial dimensions of images by scaling them to a uniform resolution, enabling consistent input during the training process. In this study, all images were resized to 640 × 640 pixels, which corresponds to the default input resolution commonly used in YOLO-based object detection models. Standardizing the image size helps maintain computational efficiency while ensuring that object features remain recognizable during model training.

Following the resizing stage, data augmentation was applied to increase the diversity of the training dataset and improve the robustness of the detection model. Data augmentation generates additional variations of existing images by applying specific transformations without altering the semantic meaning of the objects. In this study, two augmentation techniques were applied: horizontal flipping and vertical flipping. These specific transformations were chosen because the dataset primarily consists of waste objects captured under controlled indoor conditions against a uniform white background. Applying simple flips effectively introduces orientation-invariant features to the model while preserving the baseline visual characteristics and preventing the introduction of unrealistic artifacts. An illustration of the augmentation techniques used in this study is presented in Figure 2.

(a) (b)

Figure 2. Augmentation process (horizontal flip and vertical flip)

2.3. Implementation of the YOLO11 Model

You Only Look Once (YOLO) is one of the modern deep learning models that has undergone continuous development and is among the most frequently updated. YOLO11 is the latest version in the series of real-time object detection models developed by Ultralytics, setting a new standard in terms of accuracy, speed, and efficiency. Building on the advancements of its predecessors, YOLO11 introduces various improvements in both architecture and training methods, making it a flexible and reliable choice for a wide range of computer vision tasks [26].

As depicted in Figure 3, YOLO11 introduces several architectural innovations compared to its predecessors, including the use of the C3k2 block as a replacement for C2f to improve computational efficiency, as well as the integration of the C2PSA block, which serves as a spatial attention mechanism to help the model focus more effectively on important areas within an image. In addition, YOLO11 retains the SPPF (Spatial Pyramid Pooling – Fast) module to enhance multi-scale feature extraction.

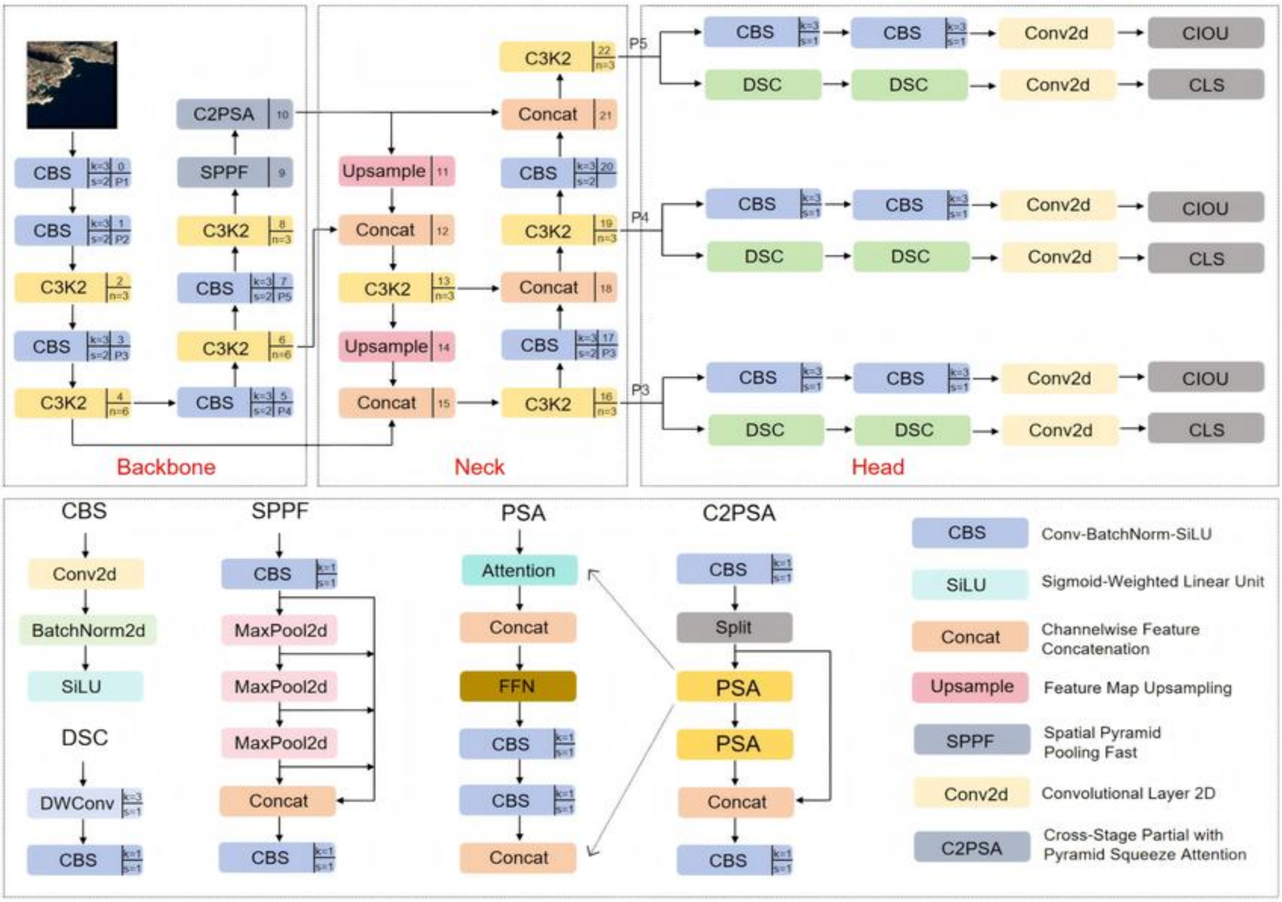


Figure 3. YOLO11 architecture (adapted from: [27])

2.4. Model Training

After completing the data preprocessing and augmentation processes, the YOLO11n model was trained on a local workstation equipped with an NVIDIA GeForce RTX 4060 Ti GPU (16 GB VRAM). In this study, the pre-trained YOLO11n.pt weight provided by Ultralytics was used as the initial model for object detection. The training process utilized the default configuration of the YOLO11 framework without modifying key hyperparameters. The model was trained using a batch size of 16 and an initial learning rate of 0.01, while the optimizer was automatically selected by the YOLO11 training framework. These baseline settings were adopted to evaluate the performance of the YOLO11n architecture for waste object detection.

2.5. Evaluation

The evaluation stage was conducted to measure the performance of the YOLO11n model in detecting and classifying six waste object categories: electronics, metal, paper, plastic, glass, and organic. Several evaluation metrics were used to assess the effectiveness of the proposed model, including confusion matrix, precision, recall, mAP, and F1-score. These metrics provide a comprehensive assessment of the model's classification and detection capabilities across different waste categories. The confusion matrix is used to analyze the distribution of correct and incorrect predictions, while precision, recall, F1-score, and mAP are used to quantify the model's overall performance in recognizing waste objects.

2.5.1. Confusion Matrix

The confusion matrix presents a detailed view of the model's performance for each class, illustrating the correspondence between the predicted and actual labels. By using a normalized confusion matrix, it becomes easier to identify both areas of high accuracy and instances of misclassification, providing valuable insights into specific classes where YOLO11 could benefit from further improvement. Together, these metrics deliver complementary perspectives on YOLO11's performance, enabling a comprehensive assessment of its detection performance [28]. As shown in Table 2, the details of the confusion matrix are presented, illustrating the distribution of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Table 2. Confusion matrix

	Actual Positive Class	**Actual Negative Class**
Predicted Positive Class	True Positive (TP)	False Positive (FP)
Predicted Negative Class	False Negative (FN)	True Negative (TN)

2.5.2. Precision and Recall

Precision represents the proportion of correctly predicted positive instances among all instances predicted as positive, indicating the model's ability to minimize false positive (FP) predictions. Recall, also referred to as sensitivity or true positive rate, measures the proportion of correctly predicted positive instances among all actual positive instances. These metrics evaluate the model's effectiveness in identifying relevant objects within the dataset. The precision and recall metrics are calculated using the following equations:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

where TP (True Positive) represents correctly detected objects, FP (False Positive) represents incorrectly predicted objects, and FN (False Negative) represents objects that were not detected by the model.

2.5.3. F1-Score

The F1-score represents the harmonic mean of precision and recall, providing a balanced measure of the model's performance by considering both false positives and false negatives. This metric is particularly useful when evaluating classification models on imbalanced datasets, where the distribution of classes is not evenly represented [28].

$$F1 - Score = 2 \times \frac{Recall\ \times Precision}{Recall + Precision} \tag{3}$$

2.5.4. Mean Average Precision (mAP)

Mean Average Precision (mAP) is a widely used evaluation metric in object detection tasks to measure the overall detection performance of a model across different classes. The metric is calculated by averaging the Average Precision (AP) values obtained for each object class. Average Precision itself is derived from the area under the precision–recall curve, which reflects the trade-off between precision and recall at different confidence thresholds. In this study, the detection performance is evaluated using mAP@0.5, where a prediction is considered correct when the Intersection over Union (IoU) between the predicted bounding box and the ground truth is greater than or equal to 0.5.

$$\mathrm{AP} = \int_0^1 P(R)dR \tag{4}$$

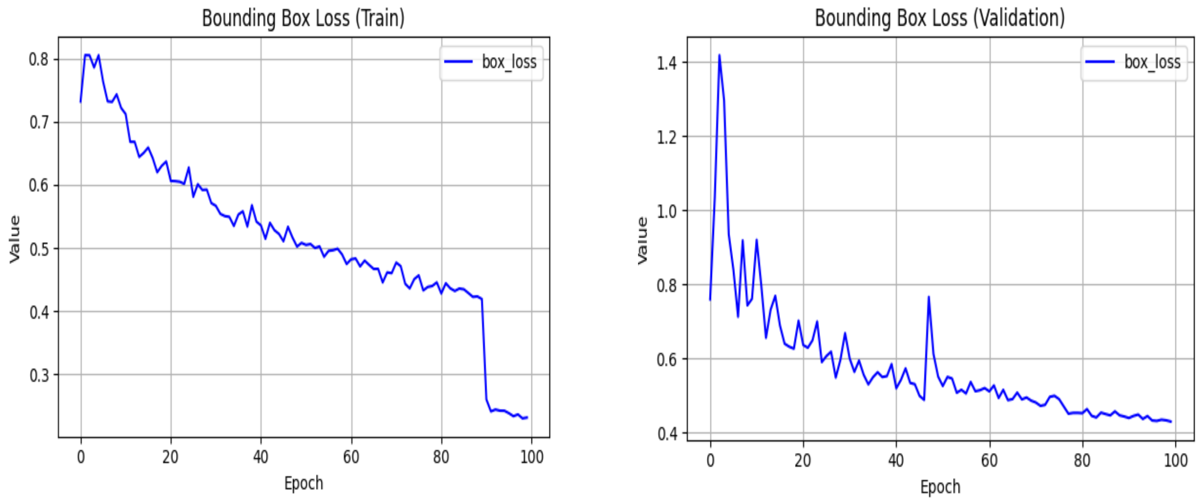
$$mAP = \frac{1}{N}\sum_{i=1}^{6} AP_i \tag{5}$$

where $P(R)$ represents precision as a function of recall, AP_i denotes the Average Precision for class i, and N is the total number of object classes evaluated.

3. RESULTS AND DISCUSSION

3.1. Model Training Performance

The YOLO11n object detection model was trained in 0.784 hours (approximately 47 minutes) using a system equipped with an NVIDIA RTX 4060 Ti GPU with 16 GB of VRAM. The model was trained for 100 epochs, during which performance evaluation was conducted on the validation dataset at the end of each epoch. This validation process was used to monitor the learning progress of the model, reduce the risk of overfitting, and ensure better generalization capability. The training performance of the YOLO11n model is illustrated in Figure 4, which presents the progression of key training metrics throughout the training process.



(a) (b)

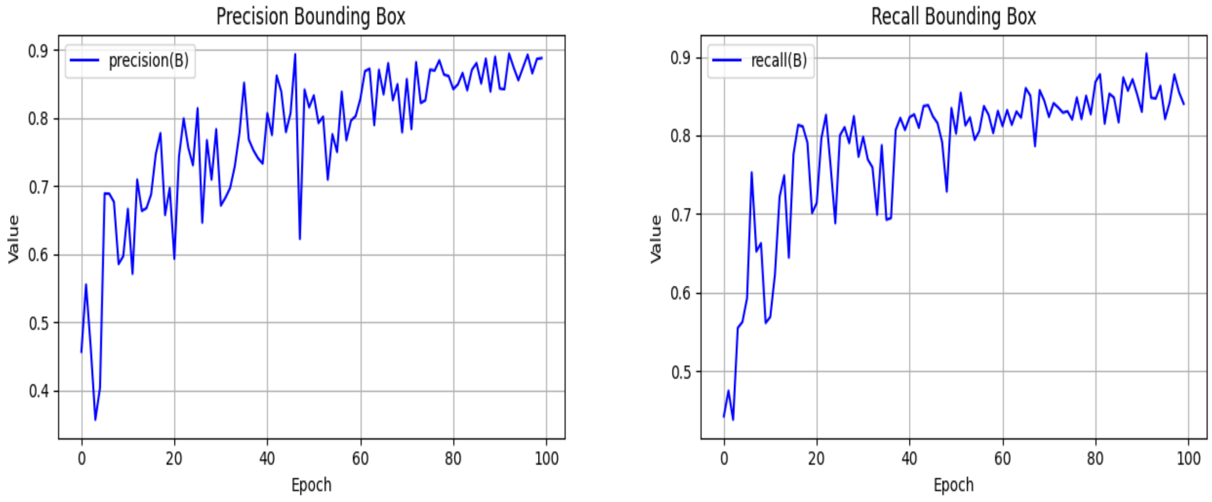
Figure 4. The results of the YOLO model training process: (a) Train and (b) Validation of box loss

Figure 4 (a) shows the trend of box_loss during the training process of the YOLO11n model. At the early stage of training, the box_loss value was relatively high (approximately 0.8), indicating that the predicted bounding boxes were still inaccurate in localizing the waste objects. As the training progressed, the box_loss value steadily decreased and reached approximately 0.4 at the end of the training process. This consistent reduction indicates that the model gradually improved its ability to accurately localize objects within the images.

Figure 4 (b) presents the progression of validation box_loss (val/box_loss), which reflects the model's localization performance on unseen validation data. Similar to the training loss, the val/box_loss value decreased from approximately 0.8 to around 0.4 as the training epochs increased. The similar decreasing trends observed in both training and validation loss suggest that the model learned stable feature representations without significant overfitting during the training process.

Figure 5 presents the evolution of precision and recall values during the training process of the YOLO11n object detection model. The precision value initially started at approximately 0.40 and gradually increased to nearly 0.90 as training progressed, indicating a significant improvement in the model's ability to correctly identify positive predictions while reducing false positives. Meanwhile, the recall value increased from around 0.50 to above 0.85, demonstrating the model's improved capability to detect the majority of target objects within the images. The consistent upward trends observed in both precision and recall indicate stable learning behavior and suggest that the model progressively improved its detection performance throughout the training process.

Figure 6 illustrates the precision–recall (PR) curves for each waste category evaluated during the training stage. The PR curves indicate that most classes achieved high detection performance, with precision and recall values consistently approaching 1.0 across a wide range of confidence thresholds. The overall detection performance of the YOLO11n model achieved a mean Average Precision at IoU threshold 0.5 (mAP@0.5) of 0.930 on the training dataset, indicating strong object detection capability across multiple waste categories.



(a) (b)

Figure 5. Precision (a) and recall (b) trends during YOLO model training

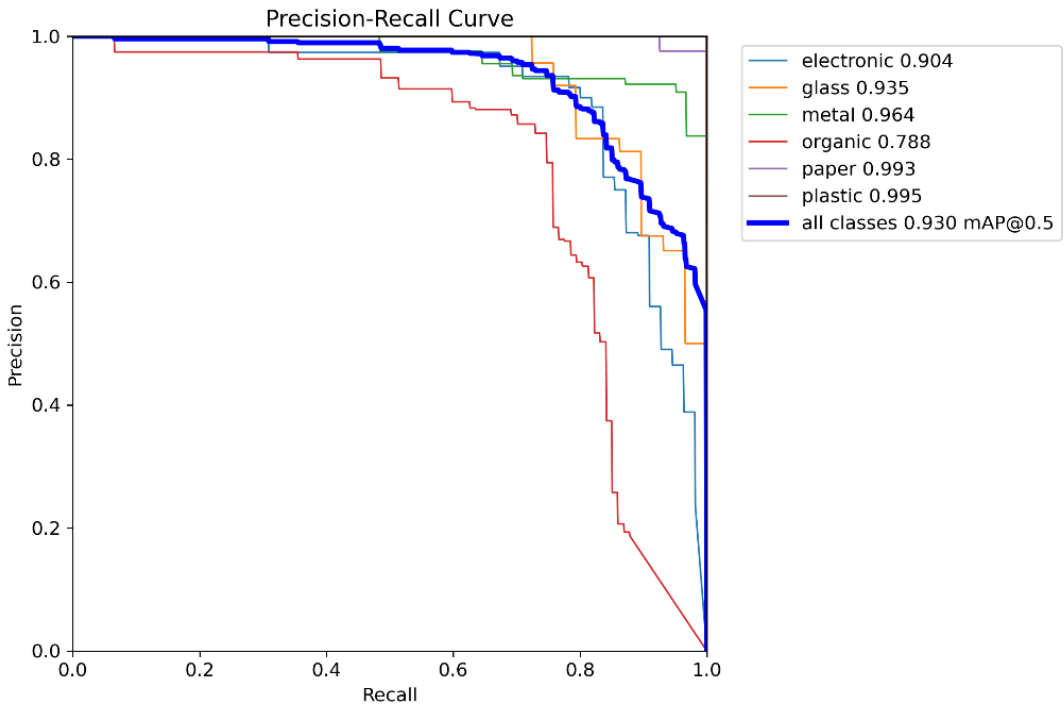


Figure 6. Precision-recall curve (mAP@0.5) during the training phase

Among the evaluated classes, Paper and Plastic achieved the highest performance, with Average Precision (AP) values of 0.995 and 0.993, respectively, indicating near-perfect detection capability. The Metal and Glass categories also demonstrated strong performance, achieving AP values of 0.964 and 0.935, respectively. The Electronics class achieved a slightly lower AP value of 0.904, while the Organic category recorded the lowest AP value of 0.788, suggesting that this class presents greater visual variability and remains more challenging for the model to detect accurately. Overall, the PR curves demonstrate that the YOLO11n model maintains a strong balance between precision and recall across most waste categories, indicating stable detection performance and reliable multi-class waste recognition capability.

Besides the Precision–Recall curve analysis, the performance of the YOLO11n model during training is further summarized using several evaluation metrics, including Precision, Recall, F1-score, mAP@0.5, and mAP@0.5–0.95, as presented in Table 3. These metrics provide a comprehensive overview of the model's detection capability across all waste categories. Precision reflects the model’s ability to minimize false positive predictions, while Recall measures the ability to correctly detect relevant objects. The F1-score provides a balanced evaluation of both metrics, and the mean Average Precision (mAP) represents the overall object detection performance at different Intersection over Union (IoU) thresholds.

Table 3. Training results

Class	Precision	Recall	F1-Score	mAP@.5	mAP@.5-.95
All	0.887	0.855	0.871	0.930	0.860
Electronic	0.923	0.782	0.846	0.931	0.801
Glass	0.784	0.897	0.837	0.962	0.869
Metal	0.923	0.772	0.841	0.980	0.954
Organic	0.865	0.701	0.774	0.771	0.645
Paper	0.975	0.979	0.977	0.995	0.986
Plastic	0.851	1.000	0.920	0.940	0.907

As shown in Table 3, the YOLO11n model achieved strong overall training performance, with Precision = 0.887, Recall = 0.855, F1-score = 0.871, and mAP@0.5 = 0.930, indicating effective object detection capability across the six waste categories. Among the evaluated classes, the Paper category achieved the highest performance, with an F1-score of 0.977, mAP@0.5 of 0.995, and mAP@0.5–0.95 of 0.986, demonstrating near-perfect detection performance. The Plastic class also showed strong results with a recall value of 1.000, indicating that all plastic objects were successfully detected by the model.

Similarly, the Metal class achieved a high mAP@0.5 of 0.980, suggesting that the model was able to effectively learn discriminative features for metallic objects. The Glass and Electronic classes also demonstrated reliable performance, with mAP@0.5 values of 0.962 and 0.931, respectively. However, the Organic class recorded the lowest detection performance, with mAP@0.5 = 0.771 and mAP@0.5–0.95 = 0.645, indicating that this category remains more challenging for the model due to higher visual variability and potential similarities with other waste types.

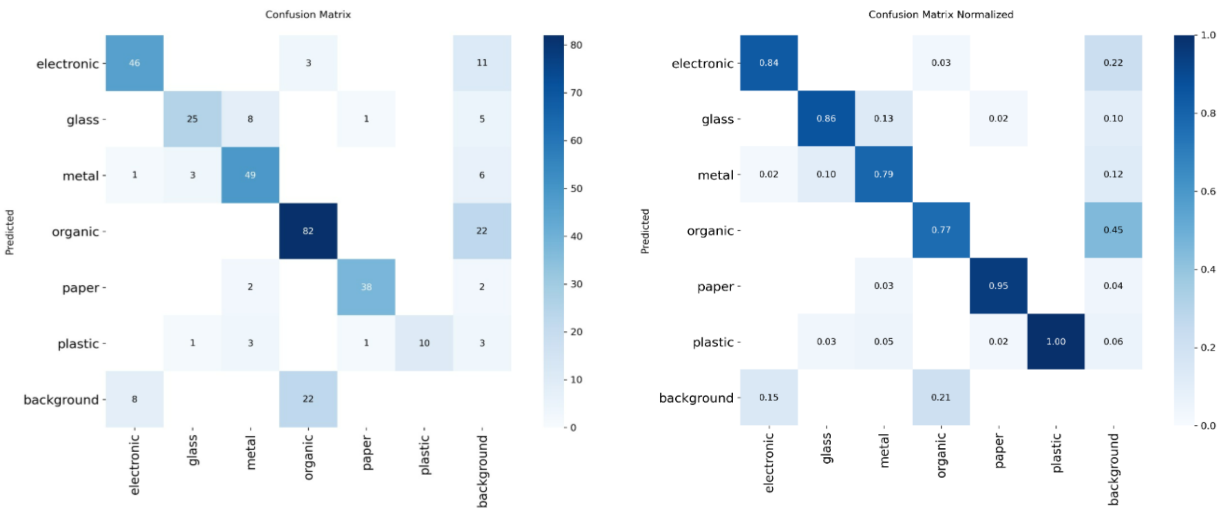
The performance of the proposed YOLO11n-based model is superior to that reported by Rahmatulloh et al. [18]. In their study, the YOLOv7-tiny based model achieved an overall precision of 0.801, recall of 0.873, and mAP@0.5 of 0.868 during training. In contrast, the model developed in this study achieved higher overall performance with precision of 0.887, recall of 0.855, and mAP@0.5 of 0.930, indicating a notable improvement in detection accuracy. A class-wise comparison further highlights the performance improvements. For instance, the glass category in Rahmatulloh et al. [18] achieved an mAP@0.5 of 0.888, whereas the proposed model obtained 0.962, showing a significant increase in detection capability. Similarly, the metal class improved from 0.921 to 0.980 mAP@0.5, while the plastic class increased slightly from 0.938 to 0.940. The paper class also demonstrated improved performance, reaching 0.995 mAP@0.5 compared with 0.970 reported in the previous study.

The most notable improvement can be observed in the organic class. In Rahmatulloh et al. [18], the organic category achieved an mAP@0.5 of only 0.597, indicating considerable difficulty in accurately detecting this class. In contrast, the proposed YOLO11n model improved the performance to 0.771 mAP@0.5, demonstrating a substantial gain in detection capability for this challenging category. These improvements can be attributed to the architectural advancements introduced in YOLO11, including improved feature extraction mechanisms and more efficient detection heads, which enable the model to capture richer semantic features from the input images. Overall, the training evaluation results demonstrate that the YOLO11n model is capable of achieving strong detection performance across most waste categories while highlighting certain classes that require further improvement.

3.2. Model Evaluation

The performance of the trained YOLO11n model was evaluated using the test dataset, which represents 10% of the total dataset following the data splitting process described in the methodology section. This subset contains images that were not used during the training or validation stages, allowing an unbiased assessment of the model's generalization capability. The evaluation metrics include precision, recall, dan F1-score. The confusion matrix provides a detailed view of the model's classification performance for each class by illustrating the relationship between the predicted labels and the corresponding ground-truth labels. In this study, a normalized confusion matrix is used, where the values represent the proportion of predictions for each class rather than absolute counts. This representation allows clearer interpretation of the classification performance and helps identify both correctly classified instances and potential misclassifications across different waste categories.

By examining the normalized values, it becomes easier to identify categories with strong detection performance as well as classes that may still require improvement. These results provide valuable insights into the strengths and limitations of the YOLO11n model for waste detection. Figure 7 presents the original and normalized confusion matrices of the YOLO11n model obtained from the test dataset. In this representation, the values indicate the proportion of predictions for each class rather than the absolute number of samples, allowing clearer interpretation of classification performance across categories. The normalized confusion matrix illustrates the distribution of predicted classes relative to the ground-truth labels, providing detailed insight into the model's classification behavior for each waste category.



(a) (b)

Figure 7. (a) Original and (b) Normalized confusion matrices on the test dataset

Based on the confusion matrix, the model achieved a Recall of 84%, 86%, 79%, 77%, 95%, and 100% for the electronic, glass, metal, organic, paper, and plastic categories, respectively. Among these classes, Paper and Plastic demonstrate the highest classification accuracy, with nearly all instances correctly predicted. This suggests that these categories possess distinctive visual features that can be effectively captured by the model. In contrast, the Organic class shows the lowest prediction accuracy and the highest level of misclassification. Several organic waste instances were incorrectly predicted as Electronic or missed entirely as Background, indicating potential visual similarities such as texture, color patterns, or object shapes between these categories and the ambient environment.

To provide a more comprehensive evaluation, a deeper multi-indicator analysis incorporating Precision and F1-score was conducted based on the raw sample counts from Figure 7. The proposed YOLO11n model achieved a Precision of 77% for electronic, 64% for glass, 83% for metal, 79% for organic, 90% for paper, and 56% for plastic. The corresponding F1-scores were calculated at 0.80, 0.73, 0.81, 0.78, 0.93, and 0.71, respectively. This cross-indicator analysis reveals a notable performance trade-off, particularly in the Plastic category. While Plastic achieved a flawless Recall of 100% (successfully identifying all 10 true plastic samples), it registered the lowest Precision (56%). The raw matrix indicates that this drop was caused by 8 false positives, where items from other categories, such as metal and glass, were incorrectly predicted as plastic due to overlapping reflective surface properties. Conversely, Paper maintains the most balanced and robust performance with an F1-score of 0.93, showing both high Precision (90%) and Recall (95%). Across all categories, the model achieved a solid macro-average Precision of 75%, an average Recall of 87%, and an overall F1-score of 0.79 on entirely unseen data. The detailed performance of YOLO11n is reported in Table 4.

Table 4. Detailed test dataset evaluation metrics for the proposed YOLO11n model

Class	Precision	Recall	F1-Score
Electronic	0.77	0.84	0.80
Glass	0.64	0.86	0.73
Metal	0.83	0.79	0.81
Organic	0.79	0.77	0.78
Paper	0.90	0.95	0.93
Plastic	0.56	1.00	0.71
Average	**0.75**	**0.87**	**0.79**

This performance difference can also be interpreted from the architectural improvements introduced in YOLO11n. The model incorporates enhanced feature extraction components, including modules such as C3k2 and C2PSA blocks, which improve the network's ability to learn richer spatial and semantic representations from the input images. The C3k2 module enables more efficient feature aggregation across multiple convolutional layers, allowing the model to capture fine-grained visual patterns from objects with varying shapes and textures. Meanwhile, the C2PSA block integrates an attention mechanism that helps the network focus on more relevant spatial regions within the image, improving its ability to distinguish between visually similar objects. These architectural enhancements contribute to the overall improvement in detection performance compared with earlier YOLO-based models. However, classes with higher visual variability, such as organic waste, may still present challenges despite these improvements, highlighting the importance of more diverse datasets and richer visual representations for further improving model generalization.

To further evaluate the effectiveness of the proposed YOLO11n model, the recall performance can also be compared with the confusion matrix reported in Rahmatulloh et al. [18], which implemented the YOLOv7-tiny architecture. Table 5 summarizes the recall comparison between both models for each waste category.

Table 5. Recall Comparison Between YOLO11n and YOLOv7-tiny [18]

Class	YOLOv7-tiny [18]	YOLO11n (Proposed)
Electronic	0.87	0.84
Glass	0.68	**0.86**
Metal	**0.84**	0.79
Organic	**0.89**	0.77
Paper	**0.96**	0.95
Plastic	0.85	**1.00**
Average	0.84	**0.87**

From Table 5, several notable improvements can be observed in the proposed model. The most significant improvement occurs in the glass category, where the recall increases from 0.68 in YOLOv7-tiny to 0.86 in YOLO11n, indicating that the proposed model is considerably more effective in detecting glass waste objects. A substantial improvement is also observed in the plastic category, where the recall increases from 0.85 to 1.00, meaning that the YOLO11n model successfully detects all plastic objects in the test dataset.

Although the Electronic, Metal, and Organic classes show slightly lower recall compared to the previous model, the differences remain relatively small and are compensated by improvements in other classes. As a result, the overall detection performance of the YOLO11n model remains higher (0.87 v.s. 0.84) and demonstrates a more balanced detection capability across the evaluated waste categories. Several factors may contribute to the improved performance of the proposed model. First, YOLO11 introduces architectural improvements in feature extraction and detection heads, enabling the network to learn richer semantic representations from the input images. These improvements enhance the model's ability to capture subtle visual differences between similar waste categories, thereby improving classification reliability.

Second, the use of data augmentation during preprocessing increases the variability of the training data and helps the model generalize better to unseen images. In this study, augmentation techniques were limited to horizontal and vertical flipping to maintain consistency with the characteristics of the dataset, which mainly consists of images captured under controlled conditions with a uniform background. Nevertheless, more diverse augmentation strategies—such as rotation, brightness variation, or noise injection—could potentially improve the robustness of the detection model and will be considered in future work.

Furthermore, the Organic class remains relatively more challenging to detect due to its higher visual variability and potential similarity with other waste categories. Organic waste objects often exhibit diverse shapes, colors, and textures, which can make it more difficult for the model to learn consistent visual patterns compared with categories such as paper or plastic that have more distinctive features. Overall, the comparison indicates that the proposed YOLO11n model achieves competitive and, in several cases, superior detection performance compared with the YOLOv7-tiny model used in Rahmatulloh et al. [18], particularly in challenging categories such as glass and plastic, while maintaining strong overall classification capability across the evaluated waste categories.

4. CONCLUSION

This study explored the use of the YOLO11n model for automated waste detection using a bounding box-based object detection approach across six waste categories: plastic, paper, glass, metal, electronics, and organic. The experimental results on the unseen test dataset demonstrate that the proposed system is capable of reliably identifying different types of waste objects and distinguishing between visually similar categories, achieving a balanced macro-average precision of 0.75, a high overall recall of 0.87, and an overall F1-score of 0.79. The findings also indicate that the YOLO11n architecture provides effective feature extraction capabilities that support accurate waste classification in image-based detection tasks, highlighted by exceptional individual recall rates such as 95% for paper and a perfect 100% for plastic. From a practical perspective, the lightweight design and computational efficiency of YOLO11n make it suitable for deployment in real-time waste management systems, such as automated waste sorting platforms, smart recycling systems, and IoT-enabled smart waste bins. These systems could support more efficient waste monitoring and sorting processes by automatically identifying waste categories and assisting decision-making in recycling or waste processing pipelines.

REFERENCES

[1] X. Meng, Z. Wen, and Y. Qian, "Multi-agent based simulation for household solid waste recycling behavior," *Resour. Conserv. Recycl.*, vol. 128, pp. 535–545, Jan. 2018, doi: 10.1016/J.RESCONREC.2016.09.033.

[2] Kementerian Lingkungan Hidup dan Kehutanan (KLHK), "Sistem Informasi Pengelolaan Sampah Nasional (SIPSN)." Accessed: Jun. 11, 2026. [Online]. Available: https://sampahnasional.kemenlh.go.id/#data-section

[3] M. MacLeod, H. P. H. Arp, M. B. Tekman, and A. Jahnke, "The global threat from plastic pollution," Jul. 02, 2021, *American Association for the Advancement of Science*. doi: 10.1126/science.abg5433.

[4] C. Pothiraj *et al.*, "Vulnerability of microplastics on marine environment: A review," Nov. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.ecolind.2023.111058.

[5] T. Dalu, T. Banda, T. Mutshekwa, & Linton, F. Munyai, and R. N. Cuthbert, "Effects of urbanisation and a wastewater treatment plant on microplastic densities along a subtropical river system," 2021, doi: 10.1007/s11356-021-13185-1/Published.

[6] M. Smith, D. C. Love, C. M. Rochman, and R. A. Neff, "Microplastics in Seafood and the Implications for Human Health," Sep. 01, 2018, *Springer*. doi: 10.1007/s40572-018-0206-z.

[7] Z. Sabarinah, "The Importance of Waste Management Knowledge to Encourage Household Waste-Sorting Behaviour in Indonesia," *Int. J. Waste Resour.*, vol. 07, no. 04, 2017, doi: 10.4172/2252-5211.1000309.

[8] M. S. Rad *et al.*, "A Computer Vision System to Localize and Classify Wastes on the Streets," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10528 LNCS, pp. 195–204, 2017, doi: 10.1007/978-3-319-68345-4_18.

[9] J. Bobulski and M. Kubanek, "Deep Learning for Plastic Waste Classification System," *Applied Computational Intelligence and Soft Computing*, vol. 2021, 2021, doi: 10.1155/2021/6626948.

[10] M. Bansal, M. Kumar, and M. Kumar, "2D Object Recognition: A Comparative Analysis of SIFT, SURF and ORB Feature Descriptors," *Multimed. Tools Appl.*, vol. 80, pp. 18839–18857, 2021, doi: 10.1007/s11042-021-10646-0.

[11] A. Rahmatulloh, H. Suhendy, and R. I. Gunawan, "MikrobatX: The Use of SIFT Feature Extraction in a Deep Learning Approach for Identification and Classification of Microscopic Fragments of Medicinal Leaves," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 13, no. 4, p. 1537, 2023, doi: 10.18517/ijaseit.13.4.18436.

[12] Y. Wei, Q. Tian, J. Guo, W. Huang, and J. Cao, "Multi-vehicle Detection Algorithm through Combining Haar and HOG Features," *Math. Comput. Simul.*, vol. 155, pp. 130–145, 2019, doi: 10.1016/j.matcom.2017.12.011.

[13] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 1, pp. 20–28, 2021, doi: 10.38094/jastt20165.

[14] C. Campbell and Y. Ying, *Learning with Support Vector Machines*. Springer, 2011. doi: 10.1007/978-3-031-01552-6.

[15] M. Kraft, M. Piechocki, B. Ptak, and K. Walas, "Autonomous, Onboard Vision-Based Trash and Litter Detection in Low Altitude Aerial Images Collected by an Unmanned Aerial Vehicle," *Remote Sens. (Basel).*, vol. 13, no. 5, p. 965, 2021, doi: 10.3390/rs13050965.

[16] P. Bharati and A. Pramanik, "Deep Learning Techniques—R-CNN and Mask R-CNN: A Survey," in *Advances in Intelligent Systems and Computing*, 2020, pp. 657–668. doi: 10.1007/978-981-13-9042-5_56.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[18] A. Rahmatulloh, I. Darmawan, A. P. Aldya, and F. M. S. Nursuwars, "WasteInNet: Deep Learning Model for Real-time Identification of Various Types of Waste," *Cleaner Waste Systems*, vol. 10, Mar. 2025, doi: 10.1016/j.clwas.2024.100198.

[19] Y. Guo *et al.*, "Bi-directional YOLOv10 with average convolution for brain tumor detection in MRI," *Brain Res. Bull.*, vol. 237, p. 111780, 2026, doi: https://doi.org/10.1016/j.brainresbull.2026.111780.

[20] X. Yang *et al.*, "Multicenter Study of YOLOv9 for Automated Detection and Classification of Supraspinatus Tendon Tears on MRI," *Acad. Radiol.*, vol. 33, no. 1, pp. 157–167, 2026, doi: https://doi.org/10.1016/j.acra.2025.10.022.

[21] S. Wu, H. Chen, Y. Zhang, K. Liao, and L. Li, "Robot visual servo based on lightweight YOLO11-SMMA for Camellia oleifera fruits harvesting," *Comput. Electron. Agric.*, vol. 243, p. 111409, 2026, doi: https://doi.org/10.1016/j.compag.2026.111409.

[22] T. Jiang and Y. Zhong, "ODverse33: Is the New YOLO Version Always Better? A Multi Domain benchmark from YOLO v5 to v11," Feb. 2025, [Online]. Available: http://arxiv.org/abs/2502.14314

[23] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," Oct. 2024, [Online]. Available: http://arxiv.org/abs/2410.17725

[24] A. Sharma, V. Kumar, and L. Longchamps, "Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species," *Smart Agricultural Technology*, vol. 9, Dec. 2024, doi: 10.1016/j.atech.2024.100648.

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[26] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO11: Real-Time Object Detection and Computer Vision Platform," 2024.

[27] J. Bai, W. Zhu, Z. Nie, X. Yang, Q. Xu, and D. Li, "HFC-YOLO11: A Lightweight Model for the Accurate Recognition of Tiny Remote Sensing Targets," *Computers*, vol. 14, no. 5, 2025, doi: 10.3390/computers14050195.

[28] M. A. R. Alif, "YOLOv11 for Vehicle Detection: Advancements, Performance, and Applications in Intelligent Transportation Systems," Oct. 2024, [Online]. Available: http://arxiv.org/abs/2410.22898