International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 15, Issue 1 [January 2026] PP: 75-84

A Novel Image-Based Automatic Waste Classification System Using YOLOv11

Sy-Dung Vu, Phan-Duy-Tan Nguyen, Minh-The Do, Duc-Hieu Nguyen, Mai-Duyen Nguyen, Ngoc-Khoat Nguyen*

Faculty of Control and Automation, Electric Power University, Hanoi, Vietnam.
**Corresponding Author: khoatnn@epu.edu.vn, nkn.research@gmail.com*

ABSTRACT: *Waste management constitutes a complex and persistent challenge that requires intelligent approaches to improve operational efficiency while minimizing dependence on manual labor. Recent progress in artificial intelligence and robotic systems has facilitated the development of automated solutions for waste detection and classification with increasingly high levels of accuracy. In this study, machine learning and deep learning techniques are employed to enable real-time waste identification and categorization, with particular emphasis on the YOLOv11 object detection framework. The proposed model is based on a Convolutional Neural Network (CNN) architecture and incorporates a modern backbone network for hierarchical feature extraction, along with an optimized neck structure designed to enhance multi-scale spatial feature fusion. To improve the robustness and generalization capability of the model, the dataset includes three commonly encountered waste categories-cardboard, metal, and plastic-collected under varying environmental and illumination conditions. Experimental evaluations were conducted using the YOLOv11s variant over 40 training epochs. The results demonstrate strong detection performance, achieving a mean Average Precision (mAP@0.5) of 99.5%, while maintaining an accuracy of approximately 87% under stricter Intersection over Union (IoU) thresholds. These outcomes indicate that the proposed approach is both effective and stable, supporting its practical applicability in intelligent waste management and automated sorting systems.*

Keywords *- Waste classification; Computer vision; YOLOv11; Convolutional Neural Network (CNN).*

Date of Submission: 06-01-2026 | Date of acceptance: 16-01-2026

I. INTRODUCTION

Environmental protection represents a fundamental component of global sustainable development. Alongside rapid industrialization and modernization, societies are increasingly confronted with critical challenges related to climate change and environmental pollution, which exert profound impacts on natural ecosystems, economic stability, and overall human well-being. Among these challenges, the rapid growth of solid waste generation has emerged as a pressing global issue. Recent statistics indicate that approximately 3.5 million tons of waste are generated worldwide each day. Moreover, global municipal solid waste production is projected to increase substantially, rising from 2.1 billion tons in 2023 to nearly 3.8 billion tons by 2050 [1]. In addition to environmental consequences, waste management poses a significant economic burden. Global expenditures on waste management were estimated at 252 billion USD in 2020. When indirect costs-including environmental degradation, public health impacts, and climate-related consequences associated with inefficient waste handling-are taken into account, the total economic cost is expected to nearly double, reaching approximately 640.3 billion USD by 2050 in the absence of effective mitigation strategies [1].These trends underscore the urgent need for efficient, scientifically grounded, and sustainable waste management and classification solutions to reduce environmental impacts and support long-term societal sustainability.

In Vietnam, rapid economic growth in recent years has been accompanied by a marked increase in both municipal and industrial waste generation. Source segregation is widely acknowledged as a critical prerequisite for improving recycling efficiency and reducing negative environmental impacts. Despite the implementation of multiple initiatives-such as public awareness programs, encouragement of household-level waste separation, expansion of recycling activities, and regulatory measures aimed at limiting single-use plastics-their overall effectiveness remains limited. This gap between policy intent and practical outcomes can be largely attributed to insufficient coordination during implementation, a strong dependence on voluntary public participation, and persistent limitations in waste collection infrastructure and treatment technologies. In particular, the uneven adoption and small-scale deployment of source-segregation practices have resulted in low recycling rates, thereby intensifying operational burdens across subsequent stages of the waste management system.

Within this context, waste management and classification solutions have evolved in both scope and complexity, with computer vision increasingly recognized as a promising approach for improving sorting efficiency. By leveraging machine learning and deep learning techniques, computer vision systems enable the identification and classification of waste materials based on visual characteristics. The incorporation of image-based processing reduces dependence on manual sorting while enhancing accuracy and automation in waste recognition tasks, thereby contributing to improvements in overall system efficiency. Such approaches are particularly well suited for deployment in centralized waste treatment facilities, automated conveyor-based sorting lines, and urban pilot-scale applications. As a result, computer vision–based waste classification can be regarded as a foundational component for the future large-scale adoption of intelligent waste management systems.

In recent years, despite increasing environmental awareness in Vietnam, the adoption of Artificial Intelligence (AI)–based approaches for waste classification has remained relatively limited and continues to face several practical challenges. In contrast, international studies have extensively investigated the use of image processing and deep learning techniques for automated waste sorting applications.For example, Chan Jia Yi and Chong Fong Kim (2024) employed Convolutional Neural Networks (CNNs) for AI-driven waste classification, reporting a training accuracy of 80.66% and a validation accuracy of 77.62% [2]. Earlier, in 2022, Matko Glučina et al. developed a YOLOv4-based model for the detection of reusable packaging. By incorporating the Mish activation function and utilizing a maximum batch size of 20,000, their approach achieved a peak mean Average Precision (mAP) of 99.96% with a minimum average error of 0.3643 [3]. In 2020, Aye et al. proposed a YOLO-based neural network combined with a Variational Autoencoder (VAE) for intelligent waste detection, attaining an accuracy of 69.70% with approximately 32.1 million parameters while maintaining a processing speed of 60 frames per second (FPS) [4-6].

Overall, both domestic and international studies underscore the significant potential of leveraging computer vision and deep learning technologies for automated waste classification. This research direction not only addresses the existing gap in waste sorting automation within Vietnam but also demonstrates high practical applicability for industrial zones, educational institutions, and urban environments.

II. EXPERIMENTAL DETAILS

2.1 Overview of YOLOv11

Figure 1 presents the overall architecture of the proposed YOLOv11-based waste recognition model, which is designed to provide accurate object detection while maintaining real-time performance on computationally constrained devices. The architecture is organized into three principal components: the backbone, the neck, and the prediction head. This modular configuration supports efficient information flow and enables effective feature extraction across multiple spatial scales.

The backbone is responsible for extracting hierarchical feature representations from the input images, which have a resolution of 640×640×3. The network architecture adopts an alternating arrangement of convolutional (Conv) layers and C3k2 blocks, a key structural component introduced in YOLOv11 to enable effective learning of complex visual patterns while preserving computational efficiency. At the final stage of the backbone, a Spatial Pyramid Pooling–Fast (SPPF) module is incorporated to enlarge the receptive field. This design allows the model to capture multi-scale contextual information without changing the spatial resolution of the resulting feature maps.

The neck component performs multi-scale feature fusion through a combination of upsampling and feature concatenation operations. In this stage, a C2PSA block is incorporated, which applies a PSA-based attention mechanism to emphasize informative waste-related features while suppressing background interference. In addition, the integration of C3k2 blocks enables further refinement of feature representations, thereby improving the balance between high-level semantic information and fine-grained spatial details.

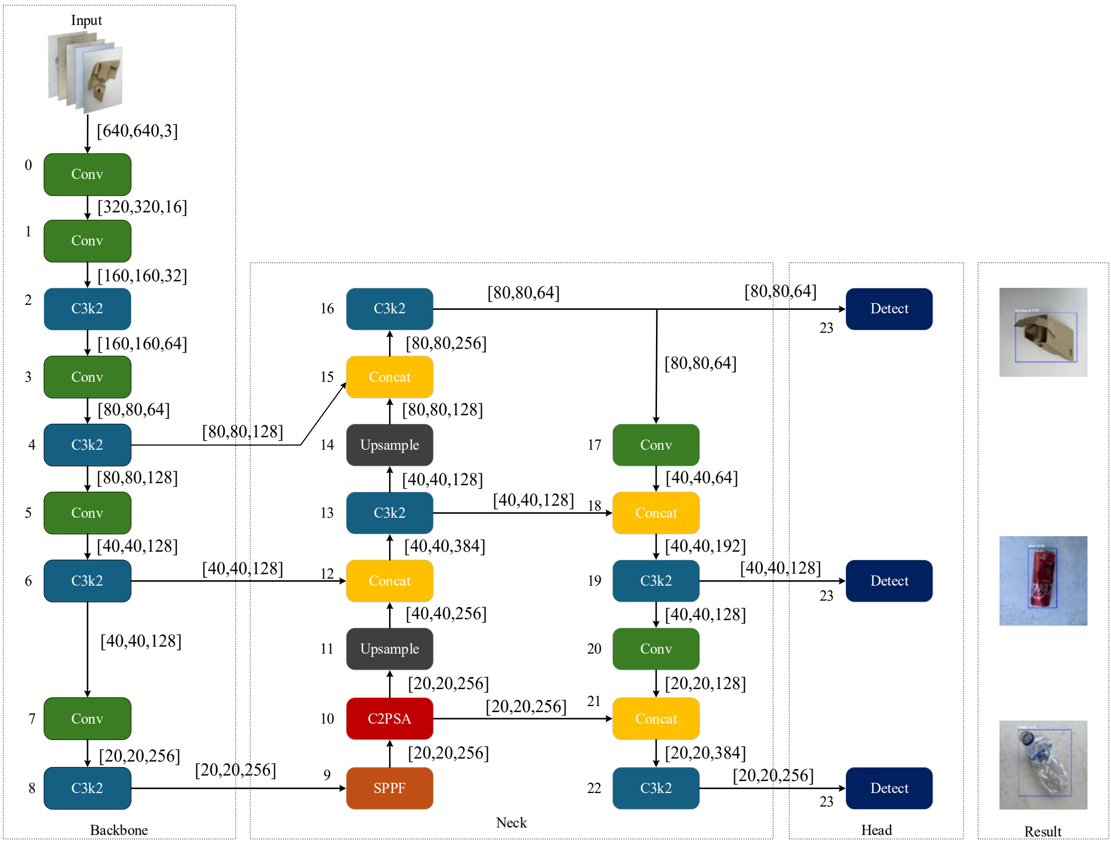

Fig. 1 The YOLOv11 architecture developed by Ultralytics

Detection Head and Empirical Findings: The prediction component utilizes a multi-head architecture with three detection layers at resolutions of 80x80, 40x40, 20x20. This configuration ensures high sensitivity to multi-scale waste objects, from fine-grained debris to bulkier items like cartons. These heads execute the computation of spatial coordinates, confidence metrics, and class probabilities. Experimental evaluations indicate that the model maintains high recognition accuracy for plastic, metal, and cardboard across varying illumination levels and perspectives.

2.2 Data Collection, Preprocessing, and Dataset Construction

Data collection is pivotal in the construction and development of a robust waste classification system. This study utilizes a dataset comprising 1,500 waste images, categorized into three distinct classes as detailed in Table 1. The data acquisition process involved two primary methods: sourcing from publicly available repositories and GitHub databases, and manual collection via original photography conducted by the research team.

Table 1: Number of Training Inputs for the Waste Classification Model

Waste name	**Cardboard**	**Metal**	**Plastic**
Amount of	500	500	500
Total	1500		

Throughout the data acquisition phase, a diverse dataset was constructed by capturing waste imagery from various perspectives, illumination conditions, and environmental settings. Each class within the dataset exhibits distinct characteristics regarding morphology, dimensions, color, and texture, as detailed in Table 2.

Table 2: Representative features of the objects

Object	Shape	Color	Structure
Cardboard	Box-shaped or flat; may be folded or torn	Light brown, red, or gray with printed text/labels	Rough or absorbent surface; box-like
Metal	Can- or container-shaped	Gray or metallic	Hard surface; may be deformed
Plastic	Bottle- or container-shaped	Predominantly transparent, blue-tinted, or stained	Smooth or flexible surface

This integrated methodology aims to augment the accuracy of the waste detection and classification system across diverse environmental conditions while simultaneously bolstering its object recognition robustness. A representative sample of the dataset acquired during the collection process is presented in Fig. 2.

Carboard Metal Plastic

Fig. 2 Representative samples from the collected waste dataset

Data preprocessing is an essential step in preparing the dataset for effective model training. In this work, data augmentation is employed to increase sample diversity and improve the representativeness of the training data. This strategy contributes to enhanced model generalization and supports more accurate and reliable waste identification and classification.

In this study, several data augmentation techniques were applied to the curated image dataset, including horizontal flipping, random rotation, and brightness adjustment. These transformations were selected to ensure that the resulting dataset more closely reflects real-world operating conditions. Through the use of data augmentation, the dataset is substantially enriched without the need for additional raw data acquisition, thereby improving the model's ability to recognize waste objects under diverse environmental and lighting conditions. Fig. 3 shows data augmentation pipeline for the waste image dataset. A range of geometric and photometric transformations-such as rotation, flipping, resizing, and brightness adjustment-are applied to simulate variations commonly encountered in real-world scenarios, contributing to improved robustness and recognition accuracy of the deep learning model.

Original Image Rotation Blurring Poor Lighting

Fig. 3 Data augmentation techniques applied to the waste dataset.

During the model training phase, standardizing all input images into a consistent format is essential to ensure a uniform data stream. By normalizing the input dimensions to 640 x 640 pixels, we enhance both training efficiency and accuracy while optimizing memory consumption for the YOLOv11 model. This specific resolution offers an optimal trade-off between preservation of visual detail and computational overhead. Subsequently, the resized images were annotated according to the three classes defined in Table 1; this ensures the model focuses exclusively on relevant objects during the learning process.

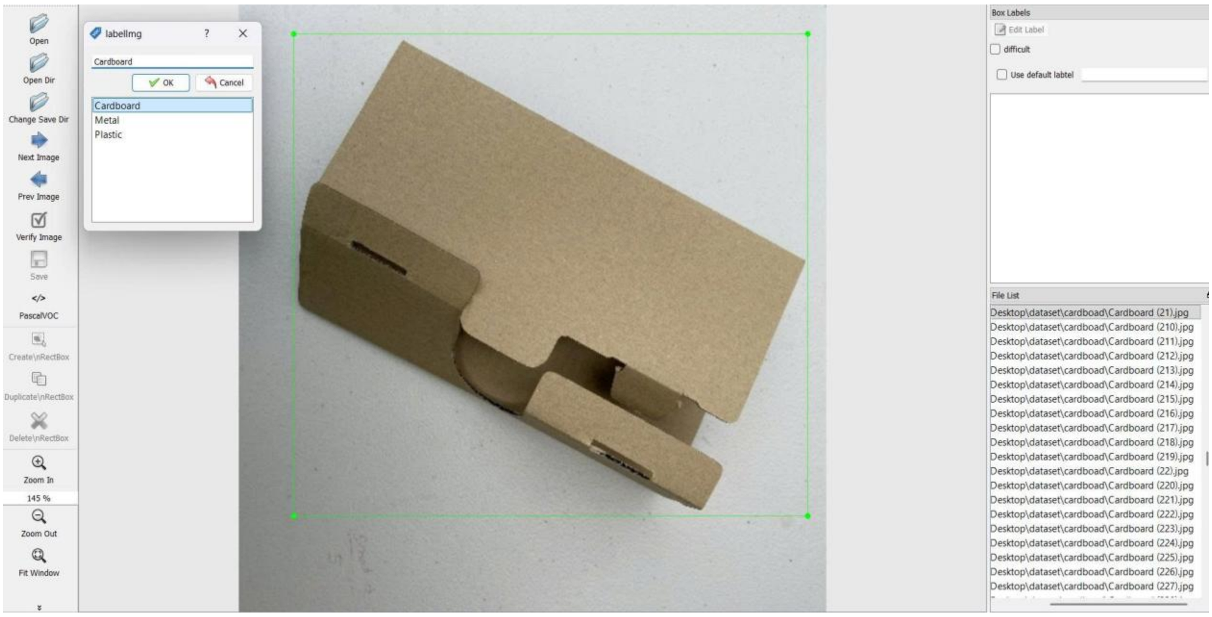

Fig. 4 Annotating waste samples using the LabelImg tool.

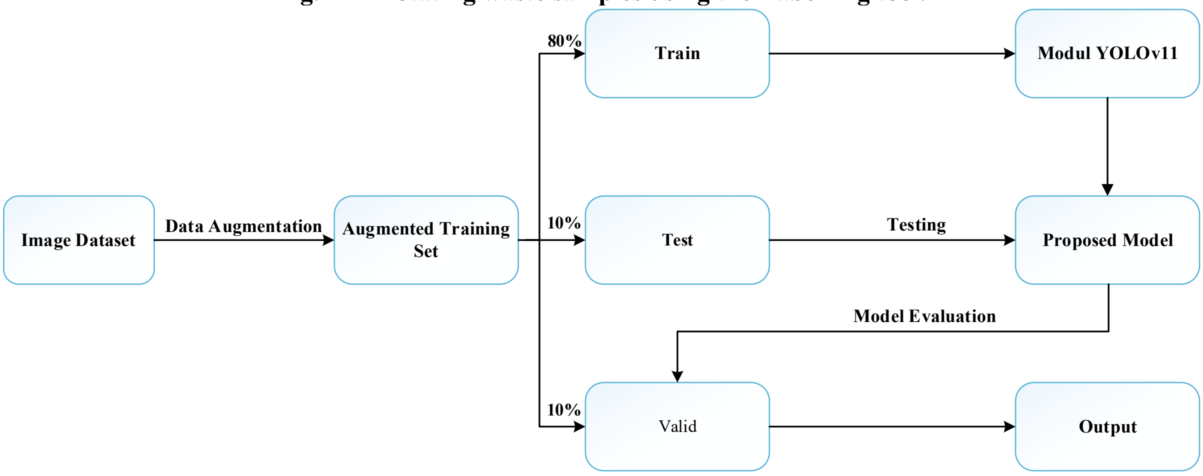

Fig. 5 Data processing workflow and dataset splitting

Following the labeling stage in Fig. 4, the dataset was divided into three subsets for training, validation, and testing. The procedural pipeline for training the custom YOLOv11s architecture is depicted in Fig. 5. To develop a scalable and reliable waste detection framework, the dataset was partitioned into three subsets: training (80%), validation (10%), and testing (10%). This data split supports effective feature learning while providing an independent validation mechanism to reduce overfitting and assess the model's ability to generalize to unseen samples. Model performance was evaluated using mean Average Precision (mAP) computed on the test set, which indicates consistent detection capability under conditions representative of real-world environments. Overall, the adopted training and evaluation pipeline contributes to improved real-time detection performance and demonstrates practical relevance for intelligent waste management applications and sustainable environmental practices.

2.3 YOLOv11s Model Training Process

Following the preprocessing stage, the data is fed into the model training phase. This study employs the YOLOv11s architecture for the object detection task. As detailed in Table 3, the YOLOv11s framework comprises three model variants, each characterized by distinct scales and levels of computational complexity.

The selection of a specific model variant involves a trade-off between inference speed and detection accuracy, which depends on application requirements and available computational resources. In addition, the choice of an appropriate variant is influenced by dataset complexity, system-level objectives, and hardware constraints. Performance evaluations consistently show a relationship between model scale and computational latency; larger models generally incur higher processing costs but provide improved detection accuracy.

Table 3: The YOLOv11 model family

Model	Size (Pixels)	mAP val 50-95	Speed of GPU ONNX (ms)	Speed of T4 TensorRT10 (ms)	Parameter (M)	Flop (B)
YOLO11n	640	39.5	56.1 ± 0.8	1.5 ± 0.0	2.6	6.5
YOLO11s	640	47.0	90.0 ± 1.2	2.5 ± 0.0	9.4	21.5
YOLO11m	640	51.5	183.2 ± 2.0	4.7 ± 0.1	20.1	68.0
YOLO11l	640	53.4	238.6 ± 1.4	6.2 ± 0.1	25.3	86.9
YOLO11x	640	54.7	462.8 ± 6.7	11.3 ± 0.2	56.9	194.9

In this study, the YOLOv11s architecture was employed for model training, with Visual Studio Code serving as the primary integrated development environment (IDE). To accelerate the training process, a workstation equipped with a 6GB GPU was utilized, leveraging parallel computing capabilities to handle the intensive workloads.

Once the development environment was configured and the data paths were verified, the training process was initiated by fine-tuning specific hyperparameters within the YOLOv11 framework. Table 4 lists the training hyperparameters, including the number of epochs, batch size, and input image size, which were tuned to optimize the learning process and ensure stable convergence with optimal predictive performance.

Table 4: Training hyperparameter configuration

Parameter	Value
Epochs	40
Batch size	12
Image size	640

III. RESULTS AND DISCUSSION

The experimental procedure was conducted utilizing the YOLOv11 model variant. The training process was executed with a batch size of 12, spanning a duration of 40 epochs. The quantitative metrics derived from this experimental phase are detailed as follows:

Table 5: Train loss

Loss type	Value
Box loss	0.49223
Classification loss	0.25872
Distribution focal loss	1.09998

Table 6: Validation losses

Loss type	Value
Box loss	0.55661
Classification loss	0.24130
Distribution focal loss	1.13871

The training and validation losses reported in Tables 5 and Table 6 demonstrate stable and consistent convergence of the YOLOv11s model. Low box and classification losses indicate effective learning of object localization and class discrimination, while the relatively higher distribution focal loss is consistent with its function in enhancing bounding box regression accuracy.

Table 7 Metrics

Metric	Value
Precision	0.99880
Recall	0.99765
mAP@50	0.99500
mAP@50–95	0.86983

As shown in Table 7, the YOLOv11s model achieved an exceptional Precision of 0.99880 and a Recall of 0.99765, demonstrating near-perfect waste detection capabilities. These metrics indicate the model’s robustness in minimizing both false positives (misidentifications) and false negatives (missed detections), thereby confirming its high efficacy and reliability for waste identification tasks.

The model achieved an mAP@0.5 of 0.99500, indicating strong performance in both detection and localization of waste objects at the standard Intersection over Union (IoU) threshold. This result reflects the effectiveness of the model in feature representation and class discrimination. When evaluated under more stringent conditions, the mAP@0.5:0.95 decreased to 0.8698. This decline suggests that performance is affected in more challenging scenarios, including cases with object occlusion, small object scales, or substantial spatial overlap among targets.

Nevertheless, the reported mAP@0.5:0.95 is considered high for practical detection scenarios. These findings indicate that the proposed YOLOv11s model achieves superior accuracy and robustness. Its balanced performance makes it an ideal candidate for real-time waste monitoring and smart environmental management applications.

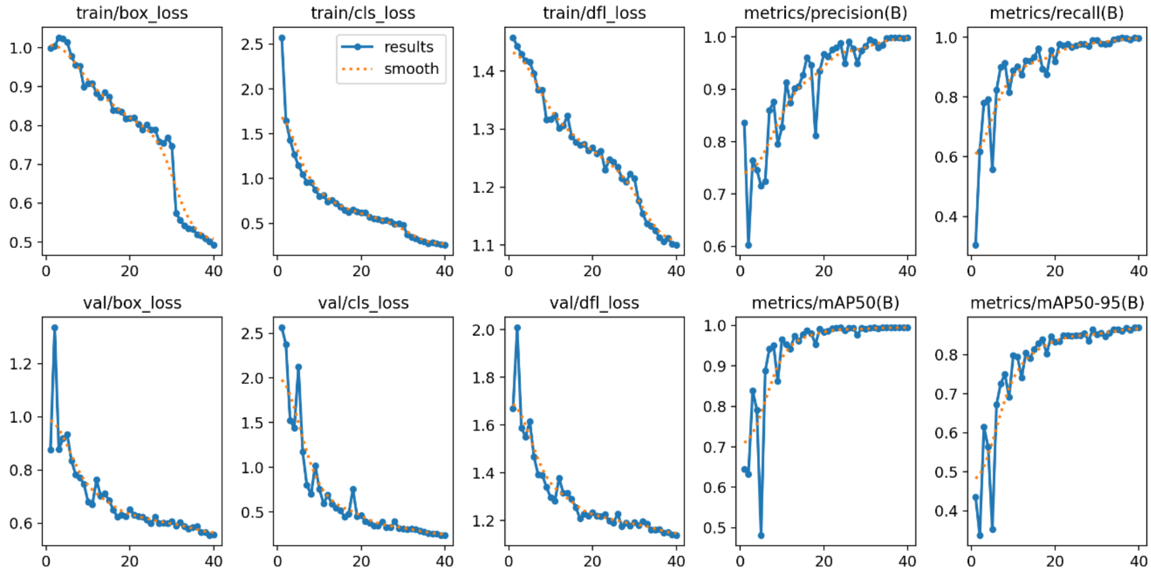

Fig. 6 Training Performance Over 40 Epochs

Analysis of the confusion matrix obtained during training showcases that the model performs reliably in distinguishing the main waste categories, namely cardboard, metal, and plastic. The metal class achieved the highest classification accuracy, with all samples correctly identified, suggesting that the model effectively captures the distinctive visual characteristics of metallic waste. For the cardboard and plastic categories, only a small number of misclassifications were observed. Overall, these results indicate that the proposed model provides accurate and consistent performance in waste classification tasks (as evidenced by the training curves in Fig. 6.

The results also reveal limitations in the model's ability to correctly classify the background class. A proportion of background samples were incorrectly labeled as cardboard or plastic, while very few background instances were accurately identified. This behavior may be attributed to class imbalance in the training data, particularly the limited number of background samples. In addition, background regions may share visual characteristics with waste objects, resulting in feature ambiguity during the classification process.

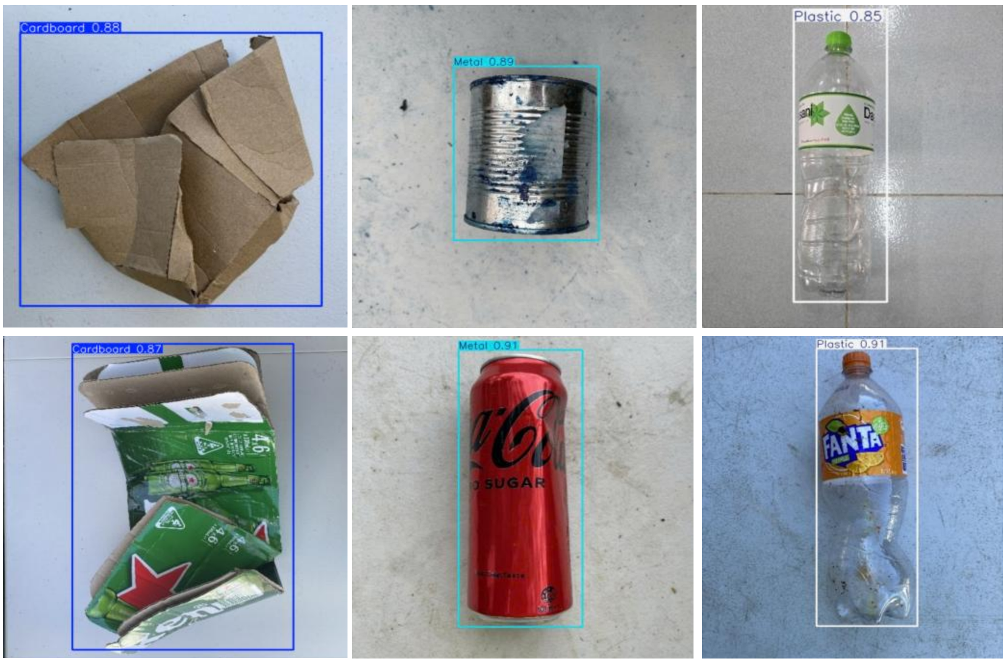

Fig. 7 Performance testing with random real-world data.

As illustrated in Fig. 7, the image-based waste classification model demonstrates stable performance and high accuracy. Objects belonging to the "Cardboard" and "Plastic" classes are correctly localized via bounding boxes and accurately labeled, with confidence scores predominantly ranging between 0.8 and 0.9. This indicates that the model not only correctly identifies the waste categories but also precisely delineates the object boundaries, even when variations in shape, size, and orientation occur. Notably, transparent plastic containers and lids are effectively detected, despite the fact that plastic features often overlap or blend with the background. The model's ability to maintain high confidence in these challenging scenarios underscores its robust feature extraction capabilities and strong generalization.

IV. CONCLUSIONS AND FUTURE WORK

This study presented a computer vision–based waste detection and classification system employing deep learning, with the YOLOv11s architecture as the core model. The system was trained using a dataset comprising three common waste categories-cardboard, metal, and plastic-collected and augmented under diverse environmental conditions. Experimental evaluations show that the proposed approach achieves strong performance, with a Precision of 0.9988, a Recall of 0.9976, and an mAP@0.5 of 99.5%. These results indicate that the model is effective and stable, supporting its practical applicability for real-time waste classification tasks.

The study also identifies several limitations, particularly in the classification of background regions, where some background samples were incorrectly labeled as waste categories. This limitation can be attributed mainly to the restricted diversity of the dataset and the overlap in visual characteristics between background regions and target objects. In addition, as the current evaluation was conducted under relatively controlled environmental conditions, further investigation is required to assess the model's performance in more complex and realistic deployment scenarios.

Future work will focus on extending the proposed system through the integration of a five-degree-of-freedom (5-DOF) robotic arm as the actuation module. This extension will enable automated manipulation tasks, including grasping, transportation, and sorting, guided by the outputs of the computer vision model. In parallel, the dataset will be expanded and rebalanced, with particular attention given to increasing the diversity of background samples and incorporating more complex environmental condition. Additionally, efforts will be directed toward optimizing the model for deployment on embedded platforms and improving synchronization between the computer vision module and the robotic control system. Collectively, these developments aim to advance the proposed framework toward a fully autonomous and intelligent waste classification system with practical applicability in industrial facilities, educational settings, and urban environments.

FUNDING STATEMENT

This research was supported by Electric Power University within the framework of the 2025 student research project, registration code ĐTNH.25/2025.

REFRENCES

[1]. LENKIEWICZ, Z. (2024). Global Waste Management Outlook 2024. Beyond an age of waste: Turning rubbish into a resource. United Nations Environment Programme. Nairobi, Quênia. DOI: https://doi.org/10.18356/765baec0-en

[2]. Yi, C. J., & Kim, C. F. (2024). AI-Powered Waste Classification Using Convolutional Neural Networks (CNNs). International Journal of Advanced Computer Science & Applications, 15(10). DOI: 10.14569/IJACSA.2024.0151009

[3]. Glučina, M., Baressi Šegota, S., Anđelić, N., & Car, Z. (2022). Automated detection and classification of returnable packaging based on YOLOV4 algorithm. Applied Sciences, 12(21), 11131. DOI: 10.3390/app122111131

[4]. Tian, M., Li, X., Kong, S., Wu, L., & Yu, J. (2022). A modified YOLOv4 detection method for a vision-based underwater garbage cleaning robot. Frontiers of Information Technology & Electronic Engineering, 23(8), 1217-1228. doi: 10.1631/FITEE.2100473

[5]. L. K. Ramadhani and B. N. Widyaningrum, "Implementation of YOLO v11 for Image-Based Litter Detection and Classification in Environmental Management Efforts," Journal of Applied Informatics and Computing, vol. 9, no. 3, pp. 617–624, Jun. 2025, doi: 10.30871/jaic.v9i3.9213.

[6]. M. Salsa Nabila, H. Herlawati, and A. Hidayat, "Pendeteksian dan Klasifikasi Sampah pada Bank Sampah Berbasis Web Menggunakan YOLOv11," Journal of Students' Research in Computer Science, vol. 6, no. 1, pp. 99–112, May 2025, doi: 10.31599/r5me0z35.

[7]. M. F. J. Permana, J. C. R. B. Gani, N. A. Fauzan, and A. Adiwilaga, "YOLOv11 Model as a Smart Solution for Waste Identification and Classification in Automated Waste Management System," Jurnal Ilmu Komputer dan Informasi, vol. 18, no. 2, pp. 207–216, Jun. 2025, doi: 10.21609/jiki.v18i2.1490.

[8]. L. A. Darisang, S. K. Dwisetio, I. S. Edbert, and A. Aulia, "Real-Time Waste Detection Using YOLO, SSD, and Faster R-CNN Integrated with CNN-Based Classification," Proc. SciTePress Int. Conf. on Computational Science and its Applications (ICCSA), 2025. (PDF).

[9]. D. Nasien, M. Hasmil Adiya, M. Farkhan, U. Sri Rahmadhani, and A. A. Samah, "Automated Waste Classification Using YOLOv11: A Deep Learning Approach for Sustainable Recycling," Journal of Applied Business and Technology, vol. 6, no. 1, pp. 68–74, 2025, doi: 10.35145/jabt.v6i1.205.

[10]. S. Kunwar, "DWaste: Greener AI for Waste Sorting using Mobile and Edge Devices," arXiv preprint, Oct. 21, 2025

[11]. Prakriti Tripathi, Theertha Biju, Maniram Thota, and Rakesh Lingam, "Image segmentation and classification of E-waste for waste segregation," arXiv preprint, Jun. 08, 2025. (áp dụng YOLOv11 vào phân loại e-waste).

[12]. Ngoc-Bao-Quang Nguyen et al., "Towards Accurate and Efficient Waste Image Classification: A Hybrid Deep Learning and Machine Learning Approach," arXiv preprint, Oct. 22, 2025. (so sánh chiến lược máy học và DL cho phân loại rác).

[13]. Yang, X., et al. (2025). Intelligent Waste Sorting for Urban Sustainability Using Deep Learning (ResNet-based). *Scientific Reports*. DOI: https://doi.org/10.1038/s41598-025-08461-w

[14]. Islam, N., Hasan Jony, M. M., et al. (2023). EWasteNet: A Two-Stream Data Efficient Image Transformer for E-Waste Classification. arXiv preprint. arXiv:2311.12823.

[15]. Fan, J., Cui, L., & Fei, S. (2023). Waste Detection System Based on Data Augmentation and YOLO_EC. Sensors, 23(7), 3646. DOI: https://doi.org/10.3390/s23073646

[16]. Kunwar, S. (2025). DWaste: Greener AI for Waste Sorting Using Mobile and Edge Devices. arXiv preprint. arXiv:2510.18513.