Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963

Human-in-the-Loop Enhanced YOLOv8 Intelligent Waste Bin for Real-Time Waste Classification

Putri Caezylia Rundupadang[1], Marvin Yonathan Hadiyanto[2*]
[1,2]Department of Electrical Engineering, Faculty of Smart Technology, Krida Wacana Christian University, Jl. Tanjung Duren Raya no. 4, Jakarta 11470, Indonesia
[1]putri.212022009@civitas.ukrida.ac.id, [2*]marvin.yonathan@ukrida.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 24-10-2026
Accepted: 10-11-2026
Published: 13-11-2026

Keywords:
Computer vision; HITL; Model improvement; Waste classification; YOLOv8n.

Brilliance: Research of Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

Waste management remains a significant environmental challenge in urban areas, where inadequate waste segregation reduces recycling efficiency and increases pollution. Although computer vision enables automated waste classification, many existing systems rely on static datasets and cannot adapt to changing conditions after deployment. This study proposes an intelligent waste bin integrating YOLOv8n with a Human-in-the-Loop (HITL) feedback framework to support real-time classification and iterative model improvement. The prototype combines a Raspberry Pi 5, Arduino Nano, USB camera, graphical user interface, and servo-based sorting mechanism to identify and separate paper, plastic, and organic waste. A dataset of 965 annotated images from campus environments and public sources was used for training and evaluation. MobileNetV2, EfficientNetB0, ResNet50, and YOLOv8n were compared in offline and real-world deployment tests. YOLOv8n showed the most robust deployment performance, achieving 75.00% accuracy, 85.19% precision, 74.44% recall, and 73.16% F1-score. HITL-based user corrections and retraining further improved classification performance. At the 10% HITL stage, offline evaluation yielded 91.19% accuracy, 91.44% precision, 91.23% recall, and 91.33% F1-score. Following HITL integration, deployment evaluation achieved 100% across these metrics on the tested samples. These findings suggest that incorporating human feedback can enhance the adaptability and practical performance of embedded waste classification systems, although broader deployment testing is needed to establish long-term reliability.

INTRODUCTION

Waste management remains a major environmental challenge in Indonesia, particularly in densely populated cities such as Jakarta (Fariz, Muis, Anggraini, Rachman, & Matsumoto, 2024). Growing urban populations and consumption have increased municipal waste generation, while the mixed disposal of organic, paper, plastic, and hazardous waste, combined with inconsistent waste separation practices, complicates sorting, reduces recycling efficiency, and increases environmental risks (Aretha Aprillia, 2021; Fasya, Ibad, El Muna, Pratiwi, & Arum, 2025). *Tempat Pengolahan Sampah Terpadu* (TPST) Bantargebang, Jakarta's primary waste disposal facility, was reported to emit up to six tons of methane per hour, while restrictions aimed at reducing incoming residual waste to 50% of its previous capacity were scheduled for August 2026 (Andre & Andreas Pamakayo, 2026; Gusti Grehenson, 2026). To reduce landfill dependence, the DKI Jakarta government has encouraged waste segregation at the source (Dinas Lingkungan Hidup Provinsi DKI Jakarta, 2026). Effective implementation, however, requires accessible sorting technologies.

Recent advances in artificial intelligence and computer vision have enabled automated waste classification at the point of disposal (Alourani, Ashraf, & Aloraini, 2025). Object detection models, particularly the YOLO family, support real-time waste identification on embedded platforms (Maity, 2023; Rathod, Rana, & Mehta, 2025). Automated sorting can improve recycling efficiency, reduce operational costs, and limit environmental pollution (Farshadfar, Khajavi, Mucha, & Tanskanen, 2025; Manik, Berawi, Gunawan, & Sari, 2024). However, reliable deployment requires models to perform under varying lighting, object orientations, backgrounds, and partial occlusion rather than only under controlled testing conditions (Wang et al., 2026).

Previous research has investigated intelligent waste sorting with deep learning and embedded hardware (Ahmad et al., 2025). Smart waste bins commonly combine cameras, Raspberry Pi boards, microcontrollers, and servo motors (Fernando & Utaminingrum, 2024). Classification approaches include feature-fusion techniques (Mohammed et al., 2023), VGG16 (Cheema, Hannan, & Pires, 2022), and YOLOv8 (Rajeev et al., 2025; Shaharia Hossen et al., 2025). Internet of Things (IoT) technologies have also been integrated for monitoring and data communication (Ahmed, Kumar Dubey, Kumar, & Dubey, 2024; Sosunova & Porras, 2022). Despite these developments, many systems rely on

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963

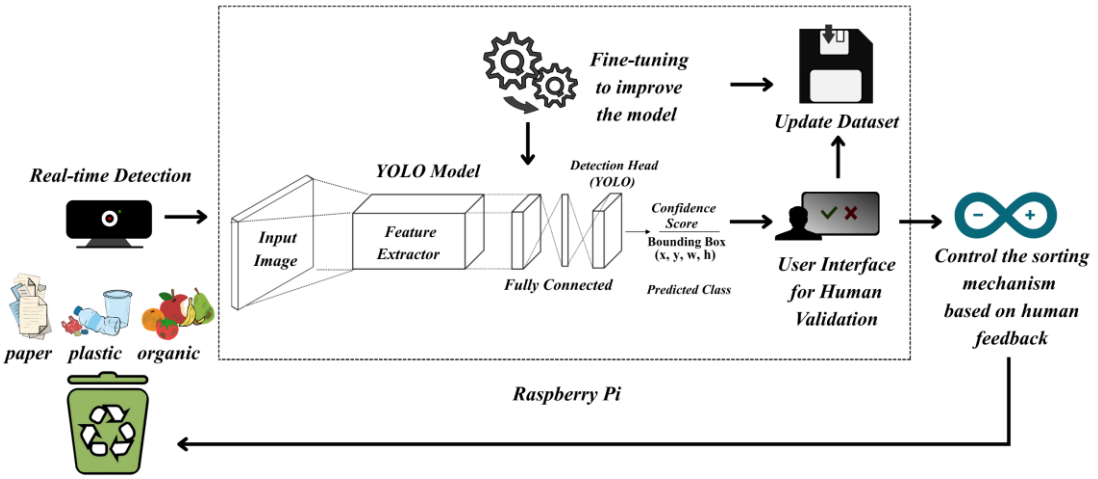
static training datasets and fixed model parameters, limiting their ability to adapt to changing real-world conditions. Furthermore, incorrect predictions are rarely utilized for model improvement, highlighting the need for adaptive learning mechanisms that incorporate feedback collected during deployment.

To address this gap, this study proposes an intelligent waste bin that integrates YOLOv8 with a HITL feedback mechanism (Lazaros, Vrahatis, & Kotsiantis, 2026). Users can verify classification results and correct misclassified predictions through a graphical user interface. Verified samples are saved and incorporated into subsequent model fine-tuning, allowing the system to learn from deployment-specific data (Kasturi, Shenfield, Roast, Le Page, & Broome, 2024). This approach combines automated sorting with human feedback to improve model adaptability under changing operating conditions.

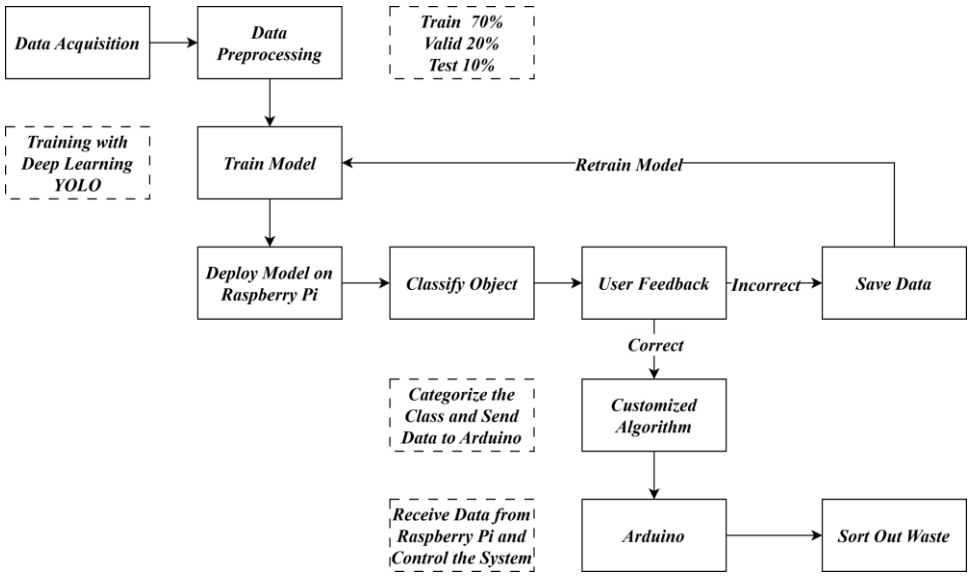
This study makes three main contributions. First, it develops an embedded YOLOv8n-based waste bin for real-time detection, classification, and mechanical sorting. Second, it implements HITL feedback to collect corrections and update the deployed model. Third, it compares classification models under offline and real-time deployment conditions and assesses how HITL adaptation affects classification performance. The findings indicate that YOLOv8n provides relatively robust deployment performance and that user-validated feedback can further improve classification reliability. The remainder of this paper is organized as follows. The METHOD section describes the system design, model development, and HITL procedure. The RESULT section presents the experimental findings, while the DISCUSSION section analyzes model performance and HITL adaptation. Finally, the CONCLUSION section summarizes the findings and suggests directions for future research.

METHOD

System Architecture



(a)



(b)

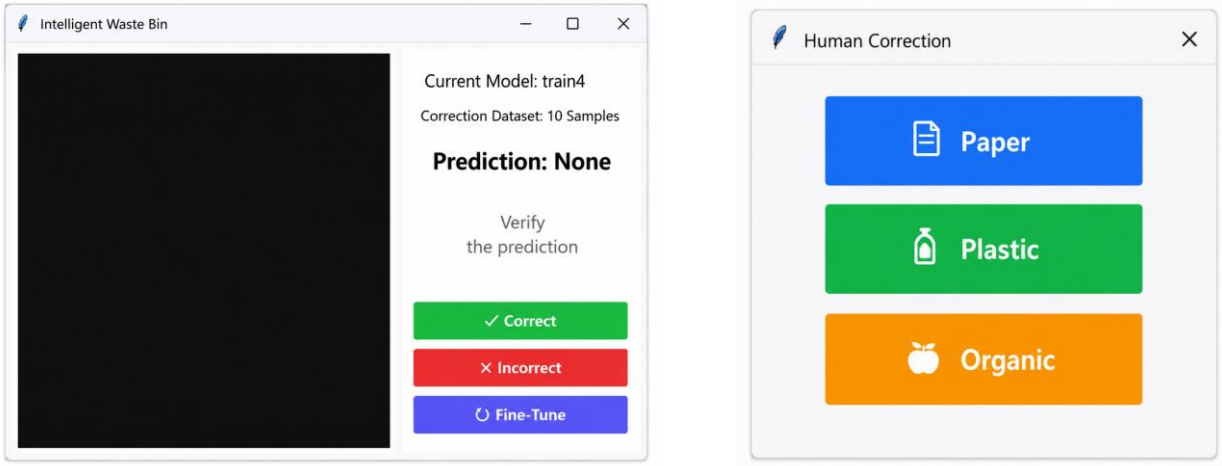
Figure 1. Overview of the proposed intelligent waste sorting system: (a) system architecture and (b) process flow diagram

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963

The proposed system integrates real-time object classification, HITL feedback, and an automatic waste sorting mechanism into a single integrated platform, as illustrated in Figure 1 (a). The system operation begins when a user places waste on the upper section of the intelligent waste bin, which serves as the object detection area. A camera continuously captures images of the waste and transmits them to the YOLOv8n model deployed on a Raspberry Pi for real-time object detection and classification. The resulting classification output is subsequently displayed on the graphical user interface (GUI), allowing users to validate the prediction results. The validated feedback data are collected and stored for further fine-tuning once the accumulated data reaches a predefined batch size.

The detailed workflow of the proposed system is presented in Figure 1 (b). The workflow begins with data acquisition and preprocessing stages, including image resizing, augmentation, and annotation, before the dataset is divided into training, validation, and testing sets. Subsequently, the YOLOv8n model is trained using the prepared dataset and deployed on the Raspberry Pi for real-time classification. The classification results are then processed by a customized control algorithm that maps the detected waste category and transmits the corresponding control commands to the Arduino Nano through serial communication. The Arduino Nano controls servo motors to execute sorting mechanisms that direct the waste into the appropriate compartment according to its detected category.



(a) (b)

Figure 2. GUI for user corrective feedback. (a) Real-time display of classification result and (b) user corrective feedback

To improve the system adaptability, we incorporate HITL based on user's feedback from GUI to continuously improve model classification performance. The GUI of the proposed system, shown in Figure 2, displays the live camera input, model information, HITL information, and prediction results. Users are provided with validation options to determine whether the classification result is correct or incorrect. If the prediction result is correct, the sorting mechanism proceeds and the waste is directed into the corresponding compartment, conversely, if the prediction result is incorrect, the user can provide corrective feedback. The corrected data are subsequently stored as new dataset entries and utilized during the fine-tuning process.

Intelligent Waste Bin Design

The proposed intelligent waste bin adopts a circular structure consisting of two main sections, namely an upper detection section and a waste storage section. Figure 3 (a) illustrates side view of the intelligent trash bin. The upper section contains a mounting structure for the camera and display above the waste detection area. The camera is positioned directly above the detection area to capture waste objects from a top-down perspective, providing a consistent field of view for image acquisition and classification. Figure 3 (b) and Figure 3 (c) illustrate the internal structure of the waste storage section. The container is divided into three separate compartments using internal partitions to accommodate paper, plastic, and organic waste. A servo-driven sorting mechanism is installed at the center of the container and connected to a movable guiding plate attached beneath the lid structure. Based on the classification result generated by the vision system, the sorting mechanism directs the waste object into the corresponding compartment.

The hardware architecture of the proposed system is shown in Figure 3 (d). The system consists of a Raspberry Pi 5 (8 GB), an Arduino Nano, a V8 Webcam 1080p, and MG996 servo motors integrated into a compact embedded platform for automatic waste sorting. The camera is connected directly to the Raspberry Pi to capture real-time images of waste objects placed on the detection platform. The Raspberry Pi performs image processing and waste classification using the deployed YOLOv8 model and subsequently transmits control commands to the Arduino Nano through serial communication. The Arduino Nano generates pulse width modulation (PWM) signals to control the servo motors

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963

responsible for actuating the sorting mechanism. Separating the high-level processing tasks on the Raspberry Pi from the low-level actuator control on the Arduino Nano improves PWM signal stability and ensures reliable servo motor operation during real-time system execution.

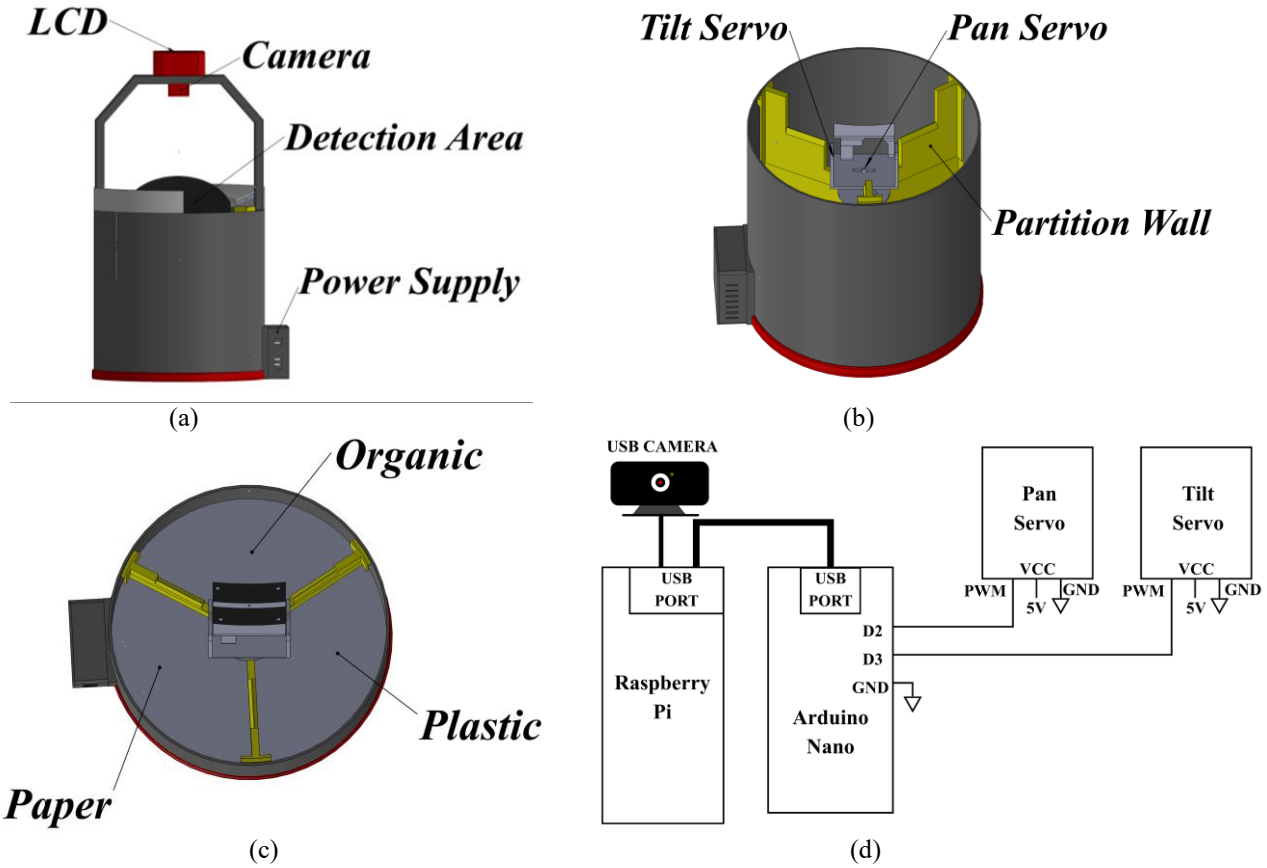


Figure 3. (a) Side view, (b) internal view, (c) internal waste compartments, and (d) hardware connection diagram of the proposed intelligent waste bin

Dataset and Model Training

The dataset used in this study was collected from two primary sources. First, waste images were captured directly from the campus environment using the camera configuration employed in the proposed intelligent waste bin, the images were collected under varying lighting conditions, object orientations, and background settings. This approach was intended to ensure that the dataset accurately represented the visual characteristics and operational conditions encountered during real-world deployment. Second, additional images were collected from publicly available online sources, including Kaggle datasets and other open-access repositories, to increase object diversity and improve model generalization. The final dataset consisted of three waste categories: paper, plastic, and organic waste. Representative objects included office paper, cardboard packaging, plastic bottles, plastic bags, food waste, fruit peels, and eggshells. The sample images from each category are presented in Figure 4.

All images were manually annotated using the Roboflow platform to generate bounding-box labels compatible with the YOLO object detection format. Following annotation, the dataset underwent preprocessing procedures including automatic orientation correction, image resizing, and format conversion. All images were standardized to a resolution of 640 × 640 pixels to ensure consistency across the dataset. The annotated dataset was subsequently divided into training, validation, and testing subsets at a ratio of 70%, 20%, and 10%, respectively. Model development was performed using Python 3.12 and the Ultralytics framework (version 8.4.41) within the Google Colab environment. A pretrained YOLOv8n model was selected as the base architecture due to its lightweight design and suitability for deployment on resource-constrained embedded platforms. Training was conducted using PyTorch 2.10.0 with an Intel Xeon CPU 2.20 GHz. The primary training hyperparameters consisted of 30 epochs, a batch size of 16, and an input image size of 320 pixels. The reduced input resolution was selected to balance detection performance and computational efficiency for subsequent deployment on the Raspberry Pi 5. After training was completed, the resulting model was exported and deployed on the Raspberry Pi 5 for real-time waste classification and sorting.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963



Figure 4. Representative samples of the three waste categories in the dataset: paper (279 images), plastic (462 images), and organic waste (224 images), the dataset comprises a total of 965 images

HITL Procedure

The HITL workflow implemented in this study is presented in Figure 5 and comprises real-time inference, user verification, correction-data acquisition, model fine-tuning, and updated-model deployment. A Tkinter-based graphical user interface displays the live camera feed, active model checkpoint, predicted waste category, and cumulative number of correction samples. For each inference result, the user validates the prediction by selecting either the correct or incorrect option. Correct predictions are directly accepted, whereas incorrect predictions trigger a pop-up interface through which the user specifies the actual category as paper, plastic, or organic waste. Each verified frame is stored in JPEG format, accompanied by a YOLO-format annotation file containing the selected class index and normalized bounding-box coordinates. Sequential file naming is applied to prevent previously collected samples from being overwritten, while a three-second interval between storage operations is enforced to minimize the acquisition of redundant or near-identical frames. Following verification, the confirmed label is transmitted to the Arduino-based actuator controller. Consequently, during HITL operation, the physical sorting destination is determined by the user-validated label rather than exclusively by the model's initial prediction, enabling the system to maintain sorting functionality while concurrently collecting deployment-specific training data.

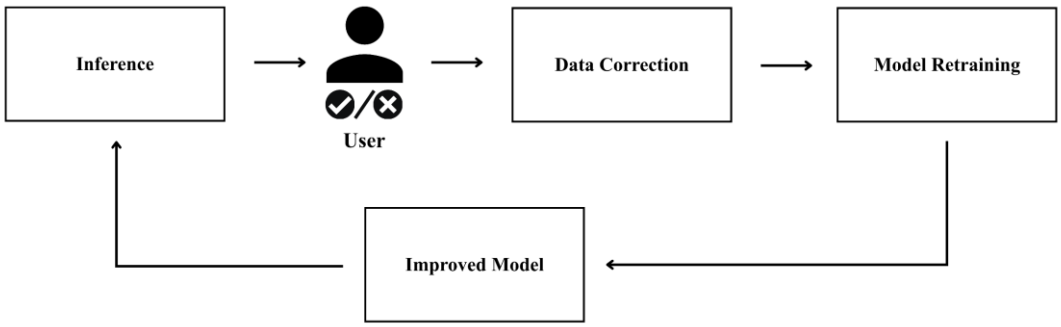


Figure 5. HITL feedback processes

Model fine-tuning is initiated manually through the GUI after a minimum of 10 feedback samples has been accumulated. The fine-tuning procedure initializes YOLOv8n using the currently active checkpoint and performs training for 20 epochs with an input resolution of 320 × 320 pixels and two data-loading workers. To preserve GUI responsiveness during training, the fine-tuning process is executed independently using multiprocessing. Upon completion, the application inspects the designated output directory for a newly generated best-performing checkpoint. Once detected, the updated model is automatically loaded and subsequently used for real-time inference without requiring the main application to be restarted. The revised model is therefore evaluated under an identical deployment configuration, including the confidence threshold, camera setup, input resolution, GUI, and actuator-control

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963

mechanism. Through this procedure, the model can progressively adapt to deployment-specific variations in illumination, background characteristics, object placement, and camera conditions. Human intervention is restricted to prediction validation, label correction, and manual initiation of fine-tuning, whereas image storage, annotation generation, model reloading, and actuator communication are performed automatically by the system.

RESULT

Implementation and Performance Analysis



(a) (b)

Figure 6. Real-time deployment of the proposed intelligent waste sorting system showing (a) the classification result and (b) the sorting mechanism

Table 1. Performance Comparison of Classification Models Under Offline and Deployment Condition

Methods	Accuracy		Precision		Recall		F1-Score	
	Offline	Deployment	Offline	Deployment	Offline	Deployment	Offline	Deployment
MobileNetV2	81.48%	37.50%	82.89%	43.59%	79.74%	37.78%	80.95%	31.48%
EfficientNetB0	97.18%	56.25%	97.79%	66.67%	96.88%	55.56%	97.32%	60.00%
ResNet50	92.25%	37.50%	92.68%	44.44%	91.76%	38.89%	92.08%	26.19%
YOLOv8n	88.42%	75.00%	89.45%	85.19%	88.08%	74.44%	88.35%	73.16%

To evaluate the classification capability and deployment feasibility of the proposed intelligent waste-sorting system, MobileNetV2, EfficientNetB0, ResNet50, and YOLOv8n were assessed under both offline and real-time deployment conditions. The offline evaluation represents model performance on a controlled test dataset, whereas the deployment evaluation is taken under the actual operation. Performance was measured using accuracy, precision, recall, and F1-score to provide a comprehensive assessment of overall classification correctness. The comparative results during offline evaluation are presented in Table 1 that reveals a considerable difference between offline performance and actual deployment performance.

During offline testing, EfficientNetB0 achieved the highest overall performance, with an accuracy of 97.18%, precision of 97.79%, recall of 96.88%, and F1-score of 97.32%. ResNet50 also demonstrated strong offline performance, achieving an accuracy of 92.25%, while YOLOv8n and MobileNetV2 achieved offline accuracy of 88.42% and 81.48%, respectively. However, the relative performance of the models changed substantially when they were deployed in the real-world intelligent waste-bin system. YOLOv8n achieved the highest deployment accuracy of 75.00%, considerably outperforming EfficientNetB0, MobileNetV2, and ResNet50, which achieved deployment accuracy of 56.25%, 37.50%, and 37.50%, respectively. YOLOv8n also obtained the best deployment precision, recall, and F1-score, reaching 85.19%, 74.44%, and 73.16%, respectively.

The magnitude of the offline-to-deployment performance gap further illustrates this difference. YOLOv8n experienced an accuracy reduction of 13.42%, from 88.42% during offline evaluation to 75.00% during deployment. In comparison, MobileNetV2 decreased by 43.98%, EfficientNetB0 by 40.93%, and ResNet50 by 54.75%. A similar pattern was observed for the F1-score. The F1-score of YOLOv8n decreased by approximately 15.19%, whereas MobileNetV2, EfficientNetB0, and ResNet50 experienced decreases of approximately 49.47%, 37.32%, and 65.89%, respectively. The deployment precision of YOLOv8n, which reached 85.19%, was higher than its deployment recall of 74.44%.

Figure 6 presents the implemented intelligent waste-bin prototype, demonstrating the integration of the deployed YOLOv8n model with the real-time classification interface and servo-driven mechanism used to direct paper, plastic, and organic waste into the appropriate compartments.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963

Performance Evaluation of HITL-Based Adaptation

Table 2. Performance Evaluation Under Variation of HITL Feedback Ratios

HITL	Accuracy		Precision		Recall		F1-Score	
	Offline	Deployment	Offline	Deployment	Offline	Deployment	Offline	Deployment
0%	88.42%	75.00%	89.45%	85.19%	88.08%	74.44%	88.35%	73.16%
2.5%	88.83%	75.00%	89.41%	85.19%	88.84%	74.44%	89.13%	73.16%
5%	89.95%	81.25%	90.22%	87.50%	90.06%	80.00%	90.14%	78.02%
7.5%	91.19%	81.25%	91.46%	87.50%	91.11%	80.00%	91.29%	78.02%
10%	91.19%	100.00%	91.44%	100.00%	91.23%	100.00%	91.33%	100.00%
12.5%	90.63%	100.00%	90.49%	100.00%	90.83%	100.00%	90.66%	100.00%
15%	90.91%	87.50%	90.96%	90.48%	92.31%	86.67%	91.63%	86.11%

The effectiveness of the proposed HITL framework was evaluated by progressively incorporating user-validated correction samples into the model fine-tuning process. The evaluation was conducted using HITL data ratios of 0%, 2.5%, 5%, 7.5%, 10%, 12.5%, and 15%, as presented in Table 2. The 0% stage represents the original YOLOv8n model before the inclusion of deployment-specific feedback, whereas the subsequent stages represent fine-tuned models using increasing proportions of user-corrected samples collected during real-time system operation. Accuracy, precision, recall, and F1-score were evaluated under both offline and deployment conditions to determine whether the feedback mechanism improved general classification capability and adaptation to the actual operating environment.

During offline evaluation, accuracy increased from 88.42% at the 0% HITL stage to 91.19% at the 10% stage. Over the same interval, precision increased from 89.45% to 91.44%, recall increased from 88.08% to 91.23%, and F1-score increased from 88.35% to 91.33%. More substantial improvements were observed during deployment evaluation. At the 0% HITL stage, the model achieved an accuracy of 75.00%, precision of 85.19%, recall of 74.44%, and F1-score of 73.16%. At the 10% HITL stage, all four metrics reached 100%.

The stage-by-stage results reveal that the effect of HITL adaptation was not strictly linear. At the 2.5% HITL stage, the offline metrics improved slightly, but the deployment results remained unchanged from the 0% stage. Deployment accuracy remained at 75.00%, while precision, recall, and F1-score remained at 85.19%, 74.44%, and 73.16%, respectively. At the 5% HITL stage, deployment accuracy increased to 81.25%, precision increased to 87.50%, recall increased to 80.00%, and F1-score increased to 78.02%. At the 7.5% HITL stage, the offline metrics continued to improve, reaching 91.19% accuracy, 91.46% precision, 91.11% recall, and 91.29% F1-score. However, the deployment metrics remained identical to those obtained at the 5% stage. The most substantial improvement occurred at the 10% HITL stage. The model achieved the best balance between offline and deployment performance at the 10% HITL stage, with an offline accuracy of 91.19%, precision of 91.44%, recall of 91.23%, and F1-score of 91.33%. More importantly, the model correctly classified all samples in the deployment evaluation set, resulting in 100% accuracy, precision, recall, and F1-score.

The 12.5% HITL stage maintained the 100% deployment score, although the offline metrics decreased slightly to 90.63% accuracy, 90.49% precision, 90.83% recall, and 90.66% F1-score. Relative to the 10% stage, these values represent decreases of 0.56% in accuracy, 0.95% in precision, 0.40% in recall, and 0.67% in F1-score. At the 15% HITL stage, performance became less consistent. Although the offline recall increased to 92.31% and the offline F1-score reached 91.63%, deployment accuracy decreased to 87.50%, precision decreased to 90.48%, recall decreased to 86.67%, and F1-score decreased to 86.11%. Compared with the initial model, the 10% HITL model therefore produced improvements of 25.00% in accuracy, 14.81% in precision, 25.56% in recall, and 26.84% in F1-score.

DISCUSSION

Model Performance and Deployment Robustness

These results indicate that all evaluated models learn the discriminative visual features from the prepared dataset, although their effectiveness varied according to model architecture and training characteristics. These findings demonstrate that the model with the highest offline performance was not necessarily the model with the highest practical deployment reliability. Therefore, although YOLOv8n did not produce the highest offline score, it exhibited the smallest degradation across the evaluated metrics and provided the most stable performance under actual operating conditions. EfficientNetB0 provides a particularly important example of the distinction between benchmark performance and deployment suitability. Although it achieved the highest offline accuracy of 97.18%, its deployment accuracy decreased to 56.25%, with a deployment F1-score of 60.00%. This result indicates that the features learned from the controlled dataset did not generalize sufficiently to the variations present during real-time operation. Similarly, ResNet50 achieved an offline accuracy exceeding 92%, but its deployment F1-score decreased to 26.19%, representing the largest performance deterioration among the evaluated models. These results emphasize that model selection for embedded computer-vision applications should not be based solely on offline accuracy.

At the aggregate evaluation level, the higher deployment precision than recall indicates that predictions generated by YOLOv8n were generally reliable when a waste category was assigned, but some waste samples remained

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963

undetected or were assigned to an incorrect category. In a practical waste-sorting application, this distinction is important because false-positive and false-negative predictions may produce different operational consequences. Incorrectly assigning a waste object to a category may cause contamination between waste compartments, while failing to correctly recognize an object may delay or prevent the sorting process. Consequently, the F1-score of 73.16% provides a more balanced representation of deployment performance than accuracy alone because it considers both prediction reliability and the model's ability to correctly recognize the evaluated waste samples.

The substantial degradation observed in MobileNetV2, EfficientNetB0, and ResNet50 suggests that the visual characteristics encountered during deployment differed from those represented in the offline test dataset. During actual operation, waste objects may appear under different illumination intensities, orientations, camera distances, viewing angles, and background conditions. Waste materials may also be folded, damaged, partially covered, contaminated, or placed away from the center of the detection area. These variations introduce a distribution shift between the training data and the deployment data, which can reduce the generalization capability of models that perform well under controlled evaluation conditions (Adhikarla et al., 2023). The comparatively robust deployment performance of YOLOv8n may be associated with its object-detection architecture.

Unlike an approach that relies only on global image-level characteristics, YOLOv8n performs localization and classification in parallel by identifying the spatial region associated with the detected waste object (Sohan, Sai Ram, & Rami Reddy, 2024). This capability may reduce the influence of irrelevant background information and allow the model to focus more directly on the visual characteristics of the waste object. The localization capability is particularly relevant in real-world operation because the object may not always be positioned consistently at the center of the camera frame. The comparison also demonstrates the importance of evaluating an intelligent waste-sorting system as an integrated cyber-physical system rather than evaluating the classification model in isolation. Real-time deployment testing provides essential evidence regarding whether a model can maintain acceptable performance after being integrated into the complete system.

Overall, YOLOv8n provided the most appropriate performance trade-off for the proposed intelligent waste bin. Its superior deployment metrics and relatively small offline-to-deployment gap indicate that it was more resilient to real-world environmental variations than the other evaluated models. Nevertheless, the deployment accuracy of 75.00% also indicates that the initial model remained vulnerable to conditions that were insufficiently represented in the original training dataset. This limitation motivated the integration of the HITL framework, through which user-validated corrections were collected as deployment-specific training samples and subsequently used for model fine-tuning.

Effects and Limitations of HITL Adaptation

The results demonstrate that the incorporation of user-corrected samples positively influenced model performance. Although these improvements were relatively moderate, they indicate that deployment-derived correction samples introduced additional information that improved the model's ability to distinguish paper, plastic, and organic waste beyond the information contained in the original training dataset. The larger increase in recall and F1-score indicates that the fine-tuned model became more effective at correctly recognizing deployment samples while maintaining reliable category predictions.

This result suggests that the initial amount of correction data was not yet sufficient to produce a measurable change in real-world classification performance. A small number of samples may have improved the model's representation of specific cases without providing enough variation to alter its predictions on the deployment evaluation set. These results represent the first clear improvement under deployment conditions and indicate that the accumulated correction samples began to capture relevant visual characteristics of the actual operating environment. Such characteristics may include the illumination level, camera viewpoint, object placement, background appearance, and physical condition of waste objects encountered by the intelligent waste bin. The unchanged deployment metrics at the 7.5% HITL stage suggest that the additional samples improved performance on the broader offline test set but did not change the predicted outcomes for the limited deployment evaluation samples. It also illustrates that improvements in offline evaluation do not necessarily produce immediate changes in deployment metrics, particularly when the deployment test set is relatively small and each individual prediction has a substantial effect on the calculated percentages. The results at the 10% HITL stage indicate that the feedback samples collected during deployment contained highly relevant information about the visual distribution of the actual system environment. By fine-tuning the model using these samples, the decision boundaries became better aligned with the conditions encountered by the camera and sorting platform.

This pattern suggests that the additional correction data remained effective for the deployment environment but did not improve generalization to the offline test set. The newly incorporated samples may have emphasized deployment-specific characteristics that were less representative of the broader original data distribution. On the deployment score, the decline from 100% scores that were obtained at the 10% and 12.5% stages demonstrates that increasing the quantity of correction data did not automatically produce continuous performance improvement. Instead, the effectiveness of fine-tuning depended on the composition, diversity, representativeness, and class distribution of the newly collected samples. Several factors may explain the non-monotonic performance observed at higher HITL ratios.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963

First, correction samples collected during consecutive system operations may contain visually similar or near-duplicate observations, such as repeated samples of similar objects, orientations, backgrounds, or illumination conditions. Such redundancy increases the size of the fine-tuning dataset without necessarily adding new information. Second, the number of correction samples for each waste category may not have been balanced. If one category generated more errors or appeared more frequently during data collection, the fine-tuning process may have overemphasized that category. This could improve recognition for frequently represented objects while reducing the decision margin for less represented categories. Consequently, the overall effect of HITL fine-tuning depends not only on the total number of feedback samples but also on their class distribution. Third, the additional samples may have introduced environmental variations that differed from both the offline test dataset and the earlier deployment evaluation conditions. Changes in lighting, object position, camera exposure, background appearance, or waste condition could alter the visual distribution of the collected samples. Fine-tuning on a relatively small and changing sample set may therefore cause the model to adapt strongly to recent observations while reducing its performance on previously learned patterns.

The decline observed at the 15% stage may also indicate the beginning of overfitting to recently collected deployment data. Because the fine-tuning procedure used a limited number of additional samples, repeated training updates could cause the model to specialize in the most recent examples. This effect may reduce generalization when subsequent evaluation samples differ from those used during fine-tuning. The results therefore demonstrate that HITL adaptation should not be interpreted as a process in which more correction data always leads to better performance. Appropriate sample selection, class balancing, diversity control, and validation are necessary to ensure that each fine-tuning stage contributes meaningful information. An important finding is that the relative improvement during deployment evaluation was substantially greater than the improvement observed during offline evaluation. The offline F1-score increased by only 2.98% between the 0% and 10% stages, whereas the deployment F1-score increased by 26.84%. This difference suggests that the HITL samples primarily addressed the distribution gap between the original dataset and the actual operating environment. The correction data captured visual characteristics that were insufficiently represented during initial model development, allowing the fine-tuned model to become more suitable for the camera configuration, illumination conditions, background, and object placement encountered in the intelligent waste bin. Nevertheless, the 100% deployment scores obtained at the 10% and 12.5% stages should be interpreted cautiously. These results indicate perfect classification on the deployment samples used in this evaluation, but they do not necessarily demonstrate perfect performance under all real-world conditions. The deployment evaluation set was limited, and a small number of samples can produce substantial changes in percentage-based metrics. Additional testing involving a larger number of objects, multiple operating sessions, different users, wider environmental variations, and balanced representation of all waste categories would be required to confirm the long-term reliability and generalizability of the adapted model.

Overall, the experimental results demonstrate that the proposed HITL framework can substantially improve the adaptability of an embedded waste-classification system by converting user corrections into deployment-specific training data. The best balance between offline performance and deployment adaptation was obtained at the 10% HITL stage. However, the performance fluctuations observed at the 12.5% and 15% stages confirm that the quality, diversity, balance, and representativeness of correction data are more important than data quantity alone. Future implementation should therefore incorporate systematic feedback-sample selection, duplicate filtering, class-balancing strategies, and validation-based model-update criteria. These measures could reduce overfitting, prevent degradation after subsequent updates, and support more stable long-term adaptation under changing real-world operating conditions.

CONCLUSION

This study developed and evaluated an adaptive intelligent waste-sorting system that integrates YOLOv8-based object detection, embedded computing, electromechanical actuation, and a HITL feedback mechanism within a unified real-time platform. The system was successfully implemented using a Raspberry Pi 5, an Arduino Nano, a USB camera, and a servo-driven sorting mechanism to detect, classify, and physically separate paper, plastic, and organic waste. Comparative evaluation of MobileNetV2, EfficientNetB0, ResNet50, and YOLOv8n demonstrated that high offline performance did not necessarily correspond to reliable performance under actual deployment conditions. Although several classification models achieved higher accuracy during offline testing, YOLOv8n exhibited the most robust real-world performance and the smallest degradation between offline and deployment evaluations. During deployment, YOLOv8n achieved an accuracy of 75.00%, precision of 85.19%, recall of 74.44%, and F1-score of 73.16%. These findings indicate that the lightweight object-detection architecture of YOLOv8n provides an appropriate balance between classification capability, localization performance, computational efficiency, and deployment robustness for resource-constrained embedded waste-sorting applications.

The integration of the HITL framework further improved the adaptability of the deployed system by converting user-validated corrections into deployment-specific training data for incremental model fine-tuning. A favorable balance between offline and deployment performance was obtained at the 10% HITL stage, with an accuracy of 91.19%, precision of 91.44%, recall of 91.23%, and F1-score of 91.33%. The fine-tuned model also achieved 100% accuracy, precision, recall, and F1-score on the deployment evaluation samples at the 10% and 12.5% HITL stages, demonstrating

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963

that feedback collected under actual operating conditions effectively captured visual characteristics that were insufficiently represented in the original dataset. Nevertheless, the perfect deployment score should be interpreted cautiously because it was obtained from a limited evaluation set and may not fully represent long-term performance across broader operational conditions. Moreover, the fluctuating results observed at higher HITL stages indicate that increasing the number of correction samples does not automatically guarantee further improvement, particularly when the newly collected data contains limited diversity, redundant observations, or imbalanced class representation. Overall, the results demonstrate that HITL-based adaptation can reduce the effects of environmental variability and data distribution shifts while maintaining real-time sorting functionality. Future work should therefore focus on expanding the dataset and waste categories, increasing the diversity and scale of deployment samples, applying systematic sample-selection and class-balancing strategies, investigating continual-learning methods that reduce overfitting and catastrophic forgetting, and conducting longer-term evaluations under varying illumination, backgrounds, object conditions, and camera configurations. These developments are expected to improve the scalability, reliability, and practical applicability of adaptive intelligent waste-management systems in real-world environments.

ACKNOWLEDGMENT

The authors wish to express their gratitude to Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Kristen Krida Wacana (UKRIDA) for providing financial support for this publication.

REFERENCES

Adhikarla, E., Zhang, K., Yu, J., Sun, L., Nicholson, J., & Davison, B. D. (2023). *Robust Computer Vision in an Ever-Changing World: A Survey of Techniques for Tackling Distribution Shifts* (Vol. 1). https://doi.org/10.48550/arXiv.2312.01540

Ahmad, G., Aleem, F. M., Alyas, T., Abbas, Q., Nawaz, W., Ghazal, T. M., … Ibrahim, A. M. (2025). Intelligent waste sorting for urban sustainability using deep learning. *Scientific Reports*, *15*(1), 27078. https://doi.org/10.1038/s41598-025-08461-w

Ahmed, K., Kumar Dubey, M., Kumar, A., & Dubey, S. (2024). Artificial intelligence and IoT driven system architecture for municipality waste management in smart cities: A review. *Measurement: Sensors*, *36*, 101395. https://doi.org/10.1016/j.measen.2024.101395

Alourani, A., Ashraf, M. U., & Aloraini, M. (2025). Smart waste management and classification system using advanced IoT and AI technologies. *PeerJ Computer Science*, *11*, e2777. https://doi.org/10.7717/peerj-cs.2777

Andre, & Andreas Pamakayo. (2026, April 14). Per Agustus 2026 Pembuangan Sampah ke TPST Bantargebang Dibatasi, Ini Mekanisme Penanganan Pemkot Jakpus. Retrieved June 9, 2026, from Pemerintah Kota Administrasi Jakarta Pusat website: https://pusat.jakarta.go.id/berita/general/perekonomian-dan-pembangunan/per-agustus-2026-pembuangan-sampah-ke-tpst-bantargebang-dibatasi--ini-mekanisme-penanganan-pemkot-jakpus-

Aretha Aprillia. (2021). Waste Management in Indonesia and Jakarta: challenges and way forward. *Proceedings of the 23rd ASEF Summer University*, 1–18.

Cheema, S. M., Hannan, A., & Pires, I. M. (2022). Smart Waste Management and Classification Systems Using Cutting Edge Approach. *Sustainability (Switzerland)*, *14*(16). https://doi.org/10.3390/su141610226

Dinas Lingkungan Hidup Provinsi DKI Jakarta. (2026, May 9). Jakarta Ajak Warga Bergerak Bersama Pilah Sampah dari Rumah untuk Masa Depan Kota. Retrieved June 9, 2026, from Dinas Lingkungan Hidup Provinsi DKI Jakarta Website website: https://lingkunganhidup.jakarta.go.id/detail-artikel/jakarta-ajak-warga-bergerak-bersama-pilah-sampah-dari-rumah-untuk-masa-depan-kota

Fariz, R. D. Al, Muis, R., Anggraini, N., Rachman, I., & Matsumoto, T. (2024). Good Environmental Governance Roles in Sustainable Solid Waste Management in Indonesia: A Review. *Journal of Community Based Environmental Engineering and Management*, *8*(8), 45–56. https://doi.org/10.23969/jcbeem.v8i1.12035

Farshadfar, Z., Khajavi, S. H., Mucha, T., & Tanskanen, K. (2025). Machine learning-based automated waste sorting in the construction industry: A comparative competitiveness case study. *Waste Management*, *194*, 77–87. https://doi.org/10.1016/j.wasman.2025.01.008

Fasya, A. H. Z., Ibad, M., El Muna, K. U. N., Pratiwi, S. F., & Arum, S. A. S. (2025, October 27). A SYSTEMATIC REVIEW OF SOLID WASTE MANAGEMENT IN INDONESIA: GENERATION, CHARACTERISTICS, TREATMENT, AND REGULATION. *Jurnal Kesehatan Lingkungan*, Vol. 17, pp. 333–342. Airlangga University Faculty of Public Health. https://doi.org/10.20473/jkl.v17i4.2025.333-342

Fernando, L. L., & Utaminingrum, F. (2024). Rancang Bangun Sistem Klasifikasi Sampah Menggunakan Yolov8 Berbasis Raspberry Pi 4. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *1*(1), 1–5. Retrieved from http://j-ptiik.ub.ac.id

Gusti Grehenson. (2026, May 18). Bantargebang Kelebihan Produksi Gas Metana, Pakar UGM Beri Saran ini. Retrieved June 9, 2026, from Universitas Gajah Mada website: https://ugm.ac.id/id/berita/bantargebang-kelebihan-produksi-gas-metana-pakar-ugm-beri-saran-ini/

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9963

Kasturi, S., Shenfield, A., Roast, C., Le Page, D., & Broome, A. (2024). Object Detection in Heritage Archives using a Human-in-Loop Concept. In Naik N, Jenkins P, Grace P, Yang L, & Prajapat S (Eds.), *Advances in Intelligent Systems and Computing* (Vol. 1453, pp. 170–181). Cham: Springer.

Lazaros, K., Vrahatis, A. G., & Kotsiantis, S. (2026, April 1). Human-in-the-Loop Artificial Intelligence: A Systematic Review of Concepts, Methods, and Applications. *Entropy*, Vol. 28. Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/e28040377

Maity, S. (2023). YOLO (YOU ONLY LOOK ONCE) ALGORITHM-BASED AUTOMATIC WASTE CLASSIFICATION SYSTEM. *JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES*, *18*(8), 25–35. https://doi.org/10.26782/jmcms.2023.08.00003

Manik, S. L. C., Berawi, M. A., Gunawan, & Sari, M. (2024). Smart Waste Management System for Smart & Sustainable City of Indonesia's New State Capital: A Literature Review. *E3S Web of Conferences*, *517*, 05021. https://doi.org/10.1051/e3sconf/202451705021

Mohammed, M. A., Abdulhasan, M. J., Kumar, N. M., Abdulkareem, K. H., Mostafa, S. A., Maashi, M. S., … Chopra, S. S. (2023). Automated waste-sorting and recycling classification using artificial neural network and features fusion: a digital-enabled circular economy vision for smart cities. *Multimedia Tools and Applications*, *82*(25), 39617–39632. https://doi.org/10.1007/s11042-021-11537-0

Rajeev, P. A., Dharewa, V., Lakshmi, D., Vishnuvarthanan, G., Giri, J., Sathish, T., & Alrashoud, M. (2025). Advancing e-waste classification with customizable YOLO based deep learning models. *Scientific Reports*, *15*(1), 18151. https://doi.org/10.1038/s41598-025-94772-x

Rathod, V. V., Rana, D. P., & Mehta, R. G. (2025). Unmanned Aerial Vehicle-Based Real-time Garbage Detection System. *Procedia Computer Science*, *258*, 2435–2443. https://doi.org/10.1016/j.procs.2025.04.506

Shaharia Hossen, M., Mosfiqur Rahman, M., Jebail Shoukhin, A., Hasan Athin, N., Chakma, A., Azim, A., & Islam, M. M. (2025). WasteSortBot: An Innovative Solution for Multi-Categorical Waste Separator Bin Using Vision-Based Control With Robot Operating System and Cloud Computing. *IEEE Access*, *13*, 183056–183074. https://doi.org/10.1109/ACCESS.2025.3623876

Sohan, M., Sai Ram, T., & Rami Reddy, Ch. V. (2024). A Review on YOLOv8 and Its Advancements. In *Data Intelligence and Cognitive Informatics* (pp. 529–545). Singapore: Springer. https://doi.org/10.1007/978-981-99-7962-2_39

Sosunova, I., & Porras, J. (2022). IoT-Enabled Smart Waste Management Systems for Smart Cities: A Systematic Review. *IEEE Access*, *10*, 73326–73363. https://doi.org/10.1109/ACCESS.2022.3188308

Wang, Z., Zhang, Z., Wang, T., Zhang, X., Ren, Z., & Zhang, Y. (2026). Garbage Recognition based on YOLOv8-seg-An Instance Segmentation Approach for Automated Waste Sorting. *International Core Journal of Engineering*, *12*(4), 2026. https://doi.org/10.6919/ICJE.202604_12(4).0029

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.