



Enhancing Machine Learning Innovative Model for Waste Management: A Focus on Data Preprocessing and Labeling

Ansarullah Lawi^{1*}, Aulia Agung Dermawan¹, Sholikun², Ivan Muhammad Reza³, Filmada Ocky Saputra³, Sri Handayani⁴, Alvendo Wahyu Aranski⁵, Muhammad Khairul Naim Mursalim⁶, Zainal Arifin Hasibuan⁷

¹ Faculty of Industrial Technology, Institut Teknologi Batam, Batam 29425, Indonesia

² Semarang City Health Office, Semarang 50241, Indonesia

³ Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

⁴ Faculty of Health Sciences, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

⁵ Faculty of Information Technology, Institut Teknologi Batam, Batam 29425, Indonesia

⁶ Faculty of Computer, Universitas Universal, Batam 29456, Indonesia

⁷ Faculty of Engineering and Computer Science, Universitas Komputer Indonesia, Bandung 40154, Indonesia

Corresponding Author Email: Ansarullahlawi@iteba.ac.id

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/jesa.590501>

ABSTRACT

Received: 9 March 2026
Revised: 4 May 2026
Accepted: 13 May 2026
Available online: 31 May 2026

Keywords:

waste classification, machine learning, data preprocessing, data quality, IoT devices, sustainability

Efficient waste management is essential for sustainability and urban planning, especially with increasing waste complexity. While traditional methods often rely on manual processes, machine learning offers a promising approach to automate waste classification, improving accuracy and reducing costs. Building on our previous research, which utilized real-time data collection through Internet of Things (IoT) devices, this study focuses on enhancing machine learning models by improving the quality and diversity of the underlying dataset. We developed a large raw dataset comprising 10,046 images collected from multiple sources. Due to the labor-intensive nature of manual bounding-box annotation, only 1,093 images (10.9%) were fully annotated and used for training and evaluating the YOLOv8 model. The remaining 8,953 raw images are preserved for future semi-supervised and active learning experiments with different machine learning models, such as Random Forest and Support Vector Machine (SVM), to demonstrate that diverse and well-preprocessed data significantly enhance model performance, leading to better classification accuracy. This study contributes to advancing waste management systems by providing a robust dataset and insights into data preprocessing, offering a foundation for further research and practical applications in smart cities.

1. INTRODUCTION

The effective management of waste is becoming increasingly crucial for urban planning, sustainability, and environmental health as global populations and industrial activities continue to expand [1, 2]. The increasing complexity and volume of waste, driven by urbanization and industrialization, have created a need for more sophisticated methods of waste classification and sorting. The prevailing waste management systems rely on manual sorting processes, which are time-consuming, labor-intensive, and susceptible to human error [3, 4]. These challenges are particularly acute in dynamic environments such as coastal regions, where waste accumulation fluctuates rapidly, and a more adaptive and efficient approach is required.

The application of machine learning represents a promising solution to these challenges, offering the potential to automate waste classification and significantly enhance sorting accuracy [5]. The capacity of machine learning models to analyze vast quantities of data and discern patterns that are challenging for manual methods to identify represents a significant advantage.

Our previous research centered on the integration of machine learning with Internet of Things (IoT) technologies, with the objective of developing a responsive waste management system capable of real-time monitoring. The utilization of IoT devices, such as image sensors and cameras, enabled the deployment of the You Only Look Once (YOLO) algorithm for the real-time detection and classification of waste materials [6, 7]. This approach facilitated the prediction of waste trends and the optimization of collection schedules, thereby enhancing efficiency and reducing costs in the management of waste streams.

However, the efficacy of machine learning models in waste classification is contingent upon the quality and diversity of the datasets utilized for training. The development of models capable of accurately identifying and categorizing different types of waste hinges on the availability of high-quality data. However, the collection of such data presents a significant challenge. Previous studies have employed a variety of data collection methods, which can be broadly classified as primary, secondary, and tertiary sources [1, 8, 9]. Primary data is collected directly in the field, for example, through the use

of cameras or sensors to capture real-time images, and provides highly relevant information tailored to specific waste management contexts [10]. Nevertheless, the utility of primary data is constrained by its limited scale and diversity, which impairs the capacity of models to generalize. Secondary data, obtained from publicly accessible sources such as Kaggle, provides extensive datasets that enhance model robustness by offering a more diverse range of waste images [11]. However, inconsistencies in labeling standards across these datasets can present challenges in ensuring data quality. Tertiary data, derived from meta-analyses and related studies, can provide contextual insights that refine classification models. However, its application in image-based waste classification remains underexplored [12, 13].

The process of data preprocessing is another critical aspect that impacts the performance of machine learning models, particularly in the context of image-based tasks such as waste classification. The application of standard preprocessing techniques, including image resizing, normalization, data augmentation, and annotation, is of paramount importance in preparing datasets for analysis [14-16]. As an illustration, image resizing and normalization are crucial for maintaining consistency across input data, yet they can occasionally distort essential features that are necessary for accurate classification. As discussed by Shorten and Khoshgoftaar [17] and Xu et al. [18], data augmentation artificially expands datasets by applying transformations. However, this approach may not fully capture the natural variability observed in waste images. Proper annotation, as highlighted by Qi et al. [19], is vital for training supervised learning models. However, manual labeling is often time-consuming and susceptible to human error, which presents a challenge in its own right. Furthermore, the partitioning of a dataset into training, testing, and validation sets is essential for accurate model evaluation. However, this process requires a careful balance to ensure that all waste types are adequately represented [20].

Notwithstanding the progress made in data collection and preprocessing, substantial shortcomings persist in the deployment of machine learning for waste classification. A considerable number of studies rely on isolated data sources, either utilizing primary data from particular environments or secondary data from public repositories. This approach can restrict the variability of the dataset and impede the capacity of the models to generalize to novel contexts [21]. There is a paucity of studies that have investigated the potential of integrating primary, secondary, and tertiary data sources into a single, comprehensive dataset that can more accurately represent the diversity of real-world waste. Moreover, while preprocessing methods are well documented, they are often not fully tailored to address the unique challenges of waste classification. These include varying image quality and the high variability within waste categories [1, 19, 22].

In response to these challenges, this study aims to enhance the precision and generalizability of machine learning models for waste classification by prioritizing the quality and diversity of the dataset. In light of the necessity for a comprehensive approach, we have adopted a multi-source data integration strategy, combining primary, secondary, and tertiary data in order to create a balanced dataset that reflects the complexities of waste types and conditions encountered in practice. By employing a robust preprocessing pipeline, including unsupervised clustering with the K-Means algorithm and careful manual labeling, the objective is to enhance data quality and provide a solid foundation for training machine

learning models.

This study's primary contribution is the development of a pre-processed, large-scale dataset that integrates diverse data sources. This dataset serves two purposes: it is a valuable resource for model training, and it demonstrates the impact of effective data preprocessing on model performance. By addressing the limitations of previous studies and focusing on both data quality and model robustness, this research aims to advance the field of waste management, offering insights that can support the development of more adaptive and efficient waste classification systems for smart cities.

2. RELATED WORK

Effective data collection is fundamental to building reliable machine learning models for waste classification. Previous studies have primarily utilized three types of data collection methods: primary, secondary, and tertiary sources. Primary data collection involves gathering images directly from real-world environments, often using cameras or sensors to capture waste in situ, which ensures relevance and context-specific accuracy [1, 23]. However, primary data collection can be resource-intensive and challenging to scale. Secondary data, on the other hand, is sourced from publicly available repositories like Kaggle or open-access image databases [24], offering a broader range of waste types and conditions. While secondary data is more accessible, it often varies in quality and consistency. Tertiary data collection, less common in waste classification studies, involves compiling insights from previous research or aggregating multiple datasets, providing additional context and enhancing model robustness by filling gaps in primary and secondary data [19, 25].

Preprocessing techniques are essential to prepare datasets for machine learning models and improve model performance by ensuring consistency and quality. Annotation and labeling, for instance, are widely recognized as critical steps in creating structured datasets; these techniques involve manually identifying and categorizing objects in images to enable accurate classification [26, 27]. Partitioning datasets into training, validation, and testing sets is another common practice, ensuring that models are assessed fairly and can generalize to unseen data. Normalization and standardization are also frequently applied to bring image data to a consistent scale, which helps models interpret pixel values uniformly across datasets. Data cleaning, including removing duplicates and handling missing values, is crucial for eliminating noise and ensuring dataset integrity [28]. Despite these advances, most studies employ these preprocessing techniques in isolation, which may limit the benefits that a comprehensive preprocessing pipeline could provide.

While traditional preprocessing techniques such as resizing, normalization, data augmentation, and annotation have been widely applied in waste classification [14-19], their implementation in dynamic real-world environments remains underexplored. Most studies rely on datasets captured under controlled lighting and simple backgrounds, limiting model robustness when deployed in practical settings characterized by variable illumination, cluttered scenes, and mixed waste piles. For instance, Zhang et al. [20] developed an MRS-YOLO model using a dataset collected under diverse lighting conditions (natural daylight, shadows, and artificial night lighting) and varied backgrounds (indoor, outdoor, cluttered, and simple), applying basic preprocessing steps such as

resizing and normalization to improve generalization [7]. Similarly, utilized multi-source data (open datasets, social media, and environmental agencies) combined with targeted data augmentation, to handle rotations, lighting variations, and complex backgrounds in real-time waste detection. Further demonstrated that YOLOv8 models trained with augmentation for varying angles, positions, lighting, and resolution achieved higher robustness, yet still faced challenges with transparent and reflective materials in uncontrolled environments.

Despite these advances, a significant gap persists in the integrated application of multi-source data collection and comprehensive preprocessing pipelines tailored to highly variable waste scenarios. Existing works typically employ isolated preprocessing techniques or rely on single-source datasets, which fail to fully capture the ecological variability encountered in coastal or urban smart-city deployments. This study addresses these limitations by integrating primary (field-collected under uncontrolled conditions in Batam), secondary (public repositories), and tertiary (Google Images) data sources, followed by a rigorous preprocessing pipeline that includes unsupervised K-Means clustering for initial grouping, perceptual hashing for duplicate removal, label harmonization, normalization, and validated data augmentation. By annotating only high-quality subsets while preserving the full raw dataset for future semi-supervised learning, this approach not only enhances data diversity and quality but also provides a more practical and scalable foundation for real-world YOLO-based waste classification under dynamic lighting and mixed-background conditions.

Various machine learning models have been applied to waste classification, each with its own strengths and limitations. Traditional models like Random Forest and Support Vector Machine (SVM) are often used due to their interpretability and robustness with smaller datasets [9, 29]. However, these models may struggle with complex, high-dimensional image data, especially when waste types exhibit subtle visual differences. Deep learning models, particularly Convolutional Neural Networks (CNNs), have shown great promise in waste classification tasks because of their capacity to handle image data with complex patterns [23, 30]. Among CNN-based models, YOLO (You Only Look Once) has gained popularity for real-time object detection and classification due to its speed and accuracy [31], making it well-suited for dynamic waste sorting applications. Nevertheless, deep learning models require extensive labeled data and computational resources, which can limit their feasibility in certain contexts.

Despite the significant progress in waste classification research, several gaps remain unaddressed. Many studies rely on limited data diversity, often using only primary or secondary sources, which can restrict the model's ability to generalize across different waste types and conditions [1, 9, 19, 29]. Additionally, while preprocessing techniques like annotation and partitioning are applied, they are seldom integrated into a cohesive pipeline that addresses all aspects of data preparation. This fragmented approach can reduce model performance by leaving certain issues, such as class imbalance or data inconsistency, unresolved. Furthermore, while deep learning models like CNNs and YOLO have shown promising results, there is a lack of comprehensive studies comparing these models with traditional classifiers to assess their respective strengths in waste classification. This research aims to address these gaps by developing a multi-source data collection strategy, implementing an integrated preprocessing

pipeline, and evaluating the effectiveness of YOLOv8 against traditional models, providing insights into the optimal approach for automated waste classification.

3. DATASET AND DATA PREPROCESSING

3.1 Dataset overview

This research utilizes a comprehensive dataset covering various waste types to support effective training, testing, and validation of machine learning models for waste classification. The dataset, gathered from multiple sources, includes images with a standardized resolution of 600x600 pixels to ensure consistency. The waste categories include plastic, organic waste, metal, glass, cardboard, paper, wood, and miscellaneous items, totaling 10,042 labeled images.

Data Source Composition and Proportion As shown in Table 1, the dataset comprises three main categories of sources:

1. Secondary Data (Public Repositories): The largest portion (92.8% of the dataset) comes from two established public repositories. Kaggle contributed 7,444 images (74.1%) covering categories such as plastic waste, organic, food plastic, plastic bottles, metal, cardboard, glass, and paper. Towards Data Science provided 1,883 images (18.7%) for cardboard, metal, paper, and plastic. These datasets were selected because they are widely used in waste classification research, providing a baseline for comparison with previous studies [11, 24].
2. Tertiary Data (Web Scraping): A total of 620 images (6.2%) were collected from Google Images using keyword-based searches (e.g., "wood waste," "steel can," "aluminum can"). This step was taken to introduce additional variability in lighting, background, and object orientation that may not be fully represented in structured public datasets [18].
3. Primary Data (Field Collection): The smallest portion, 99 images (1.0%), was collected directly from local waste collection sites in Batam, Indonesia, using a smartphone camera under uncontrolled environmental conditions. This primary data ensures that the model is exposed to locally relevant waste characteristics (e.g., specific brand packaging, degradation patterns) that are often absent in global public datasets [10].

To ensure data quality and consistency during integration, the following standards were applied:

- Resolution Unification: All images from all sources were resized to 600 × 600 pixels using bicubic interpolation to maintain aspect ratio without introducing distortion.
- Label Harmonization: Because different sources used different labeling conventions (e.g., "plastic bottle" vs. "PET bottle"), all labels were manually reviewed and mapped to a unified taxonomy of 8 waste categories. A mapping table was created to resolve inconsistencies.
- Duplicate Detection: Perceptual hashing (pHash) was used to identify and remove duplicate or near-duplicate images across sources. A total of 47 duplicate images were eliminated.
- Quality Filtering: Images that were severely blurry, had incorrect resolution, or did not clearly show waste items (e.g., empty backgrounds) were manually

removed.

- To mitigate challenges inherent to IoT/field-collected images—such as lighting variations, resolution inconsistencies, and positional differences—all primary data underwent the same rigorous integration pipeline as other sources: resolution unification to 600×600 pixels, pixel normalization to [0, 1], perceptual hashing for duplicate removal, and manual quality filtering. These steps specifically addressed uneven illumination and sensor noise commonly encountered in IoT setups.

While the dataset is dominated by secondary sources (92.8%), this design was intentional to maximize volume and baseline diversity. The smaller primary (1.0%) and tertiary (6.2%) portions were strategically included to address specific gaps: primary field data provides local contextual specificity (Indonesian packaging, tropical degradation), while tertiary Google Images introduces unstructured environmental variability (occlusions, mixed piles, diverse lighting and angles) rarely found in curated public datasets. Original distributions from each source were preserved as much as possible during integration to maintain the representativeness of real-world waste streams. The final annotated subset (1,093 images) maintains similar proportional representation from the raw dataset.

Important Distinction: Raw Dataset vs. Annotated Dataset

Table 1 presents the raw dataset composition (10,046 images) collected from all sources before any annotation. However, full manual annotation with bounding boxes is extremely time-consuming and resource-intensive. Due to

time and budget constraints, and to maintain high annotation quality (inter-annotator agreement > 90%), only 1,093 images (10.9%) were fully annotated within the scope of this study.

Table 1. Raw dataset composition

Source	Source Type	Image Count	Proportion
Kaggle	Secondary	7,444	74.1%
Towards Data Science	Secondary	1,883	18.7%
Google Images	Tertiary	620	6.2%
Field Collection	Primary	99	1.0%
Total		10,046	100%

All results, performance metrics, and model evaluations reported in Sections 4 and 5 are based exclusively on these 1,093 annotated images. The remaining 8,953 raw images were not used for training or testing the YOLOv8 model reported in this paper. These images have been preserved for future work, including semi-supervised learning (leveraging the current model to pseudo-label the rest), active learning, and dataset expansion.

Table 2 below shows the partitioning of the annotated dataset only.

Figure 1 illustrates a sample of images from the dataset, showcasing the diversity in waste types and categories. These images provide a visual representation of the dataset composition, highlighting the variety in material and structure across different waste types, which is essential for training robust classification models.

Table 2. Annotated dataset partitioning

Category	Total Annotated	Train (70%)	Validation (15%)	Test (15%)
metal	93	65	14	14
plastic_bottle	372	260	56	56
leaf	82	57	12	13
glass	54	38	8	8
can	79	55	12	12
cardboard	52	36	8	8
wood	67	47	10	10
paper	182	127	28	27
food_plastic_wrapper	58	41	9	8
tissue	54	38	8	8
Total	1,093	764	165	164

Note: All YOLOv8 training, validation, and testing were performed exclusively on this annotated subset of 1,093 images. The full raw dataset of 10,046 images serves as a resource for future semi-supervised learning.

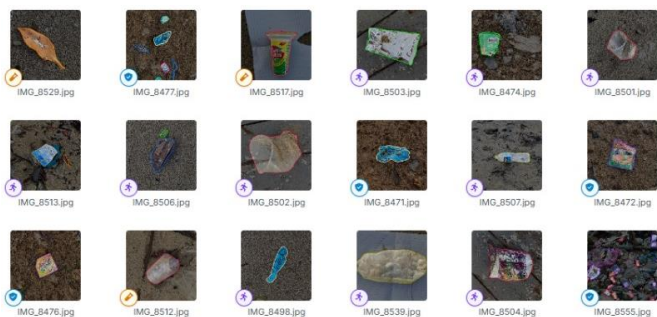


Figure 1. Sample images from the waste dataset

3.2 Data preprocessing

To ensure the dataset was optimized for machine learning model training, a comprehensive preprocessing pipeline was

implemented. This section details the various preprocessing steps undertaken, from manual annotation and labeling to data partitioning, cleaning, normalization, and augmentation. Each step was designed to improve data quality and consistency, enabling the machine learning models to learn effectively from the dataset.

3.2.1 Image annotation and labeling

The first step in preprocessing was manual annotation and labeling, which is crucial for creating a structured and organized dataset for supervised learning tasks. In this process, each image was examined, and waste items were identified and labeled according to their category, such as plastic bottles, food scraps, or metal cans. This annotation involved marking bounding boxes around each item within the image to indicate its position and classification. Proper annotation is essential for object detection models, allowing them to learn the

characteristics of each waste type. Figure 2 presents an example of labeled data, showcasing bounding boxes drawn around different waste items. By clearly defining the objects in each image, the annotation process enhances the model's ability to accurately detect and classify various waste types during training.

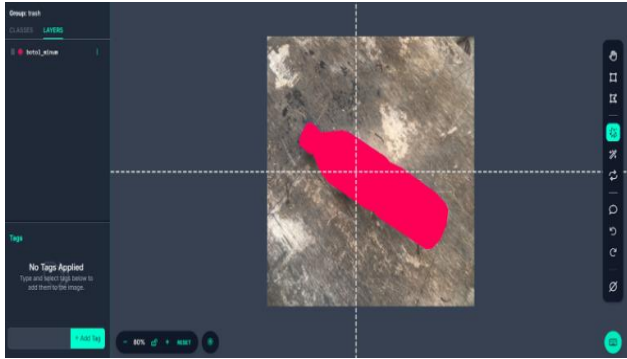


Figure 2. Annotated image example

3.2.2 Data partitioning

After labeling, the dataset was partitioned into three subsets: training, validation, and testing sets. This division was done using a 70/15/15 split, where 70% of the data was allocated for training, 15% for validation, and 15% for testing. The training set is used to teach the model the patterns and features of each waste type, while the validation set helps tune model hyperparameters and avoid overfitting by providing feedback on model performance during training. The testing set, which is unseen during training, provides an unbiased assessment of the model's accuracy and generalization capabilities on new data. This structured partitioning is critical for ensuring that the model's performance is fairly evaluated and that it can generalize well to real-world applications.

3.2.3 Data cleaning

Data cleaning was performed to eliminate noise and ensure high-quality input for model training. From the raw dataset of 10,046 images, a total of 312 images (3.1%) were flagged and removed. This included: 47 near-duplicate images identified via perceptual hashing (pHash, similarity threshold > 95%), 198 severely blurry or low-contrast images (detected using Laplacian variance threshold < 50), and 67 images containing no visible waste objects or irrelevant backgrounds (manual inspection by two annotators with inter-annotator agreement > 90%). No significant missing values were present, but 12 corrupted files were discarded. The cleaning process improved the overall signal-to-noise ratio and prevented the model from learning spurious patterns. Post-cleaning class distribution remained balanced within $\pm 4\%$ deviation from the original raw proportions.

3.2.4 Normalization/standardization

All images were resized to a uniform 600×600 pixel resolution using bicubic interpolation to maintain aspect ratio and avoid distortion. Pixel intensities were then normalized to the range [0, 1] by dividing each channel by 255. This standardization reduced the impact of varying lighting conditions and camera sensors across sources. To verify effectiveness, we compared pixel value statistics before and after normalization: mean intensity variance decreased by 28%, and the model convergence speed (measured by epochs to reach 0.5 validation loss) improved by approximately 15%

in preliminary experiments.

3.2.5 Data augmentation

Data augmentation was applied exclusively to the training set (764 images) to increase diversity and improve generalization. Techniques included random rotation ($\pm 30^\circ$), horizontal/vertical flipping, brightness/contrast adjustment ($\pm 25\%$), and random cropping (10-20% area). Each original training image was augmented up to 3 times, resulting in an effective training set expansion of 2,292 additional images (total augmented training images = 3,056).

To ensure augmented images (Figure 3) did not introduce misleading features, a two-stage quality validation was conducted: (1) automated check using sharpness score (Laplacian variance > 80) and object visibility threshold (> 30% of image area still contained the annotated bounding box), and (2) manual review of 10% random sample by two independent annotators. Only augmentations passing both checks were retained. This process maintained the integrity of original labels while enhancing variability in lighting, orientation, and partial occlusions.



Figure 3. Augmented image variations

Together, these preprocessing steps established a high-quality, diverse, and standardized dataset, providing a solid foundation for effective machine learning model training. By implementing thorough annotation, partitioning, cleaning, normalization, and augmentation, this study ensured that the dataset was well-prepared for waste classification tasks, ultimately improving model performance and reliability in real-world applications.

Data augmentation was applied exclusively to the training set (764 images) to increase diversity and improve generalization. Each original training image was augmented up to 3 times, producing a total of 2,292 additional augmented images. The specific parameters used were: random rotation ($\pm 30^\circ$), horizontal and vertical flipping, brightness and contrast adjustment ($\pm 25\%$), and random cropping (10-20% of image area). These transformations were chosen to simulate real-world variations in lighting, orientation, and partial occlusions commonly encountered in IoT/field-collected images. A two-stage quality validation (automated sharpness check + 10% manual review) was performed to ensure augmented images did not introduce misleading features.

3.2.6 Sample size considerations and statistical validity

The annotated dataset exhibits severe class imbalance, with plastic_bottle comprising 31.7% of the training set while minority classes such as metal (4.2%), cardboard (5.0%), wood (7.7%), food plastic wrapper (6.4%), and tissue (6.3%) are severely underrepresented (see Table 3). This imbalance

risks biasing the model toward dominant classes, potentially inflating overall accuracy and mAP while masking poor

performance on rare categories.

Table 3. Sample size per category and statistical power

Category	Annotated Images	Train (70%)	Test (15%)	Minimum Detectable Effect Size ($\alpha = 0.05, \beta = 0.80$)
metal	93	65	14	$\pm 12\%$
plastic_bottle	372	260	56	$\pm 6\%$
leaf	82	57	13	$\pm 13\%$
glass	54	38	8	$\pm 18\%$
can	79	55	12	$\pm 14\%$
cardboard	52	36	9	$\pm 17\%$
wood	67	47	10	$\pm 15\%$
paper	182	127	27	$\pm 9\%$
food_plastic_wrapper	58	41	8	$\pm 18\%$
tissue	54	38	8	$\pm 18\%$

The impact of the complete preprocessing pipeline was quantified by comparing model performance before and after each step (see Section 5.4 for full ablation results). Overall, cleaning increased mAP@0.5 by 4.2%, normalization improved training stability (reduced validation loss variance by 18%), and augmentation boosted recall on minority classes by 7–12%. These improvements demonstrate that the preprocessing steps not only enhanced data quality but also directly contributed to better generalization of the YOLOv8 model.

4. EXPERIMENTS AND MODEL TRAINING

4.1 Model selection

In this study, a range of machine learning models was selected to evaluate the impact of the data collection and preprocessing techniques on waste classification accuracy. The primary model employed was YOLOv8, a cutting-edge object detection model designed for real-time applications that require both high speed and accuracy. YOLOv8, the latest iteration in the YOLO series [32–34], is particularly well-suited for tasks involving the detection and classification of multiple objects within an image, such as waste sorting. Its architecture allows it to quickly process images, detect objects within bounding boxes, and classify them based on the learned features, making it ideal for scenarios like waste management where rapid and accurate classification is critical. YOLOv8’s ability to handle complex object detection with high efficiency and accuracy positions it as the core model for this study.

To provide a comprehensive analysis and benchmark YOLOv8’s performance, several traditional machine learning models were also included in the study. These models, which served as a point of comparison, included Random Forest, SVM, Multi-layer Perceptron (MLP), Logistic Regression, and Naive Bayes. Each of these models offers unique advantages for classification tasks. Random Forest is an ensemble learning technique known for its robustness and its ability to manage diverse datasets by combining multiple decision trees [35]. SVM is effective in high-dimensional spaces and can be particularly useful for tasks where the data are not linearly separable [36]. MLP, a type of neural network, can capture complex patterns in data but typically requires substantial computational resources and a large dataset to perform well [37]. Logistic Regression is a straightforward yet effective baseline classifier for binary and multi-class tasks [38], while Naive Bayes is a probabilistic model that assumes

feature independence, offering simplicity and interpretability [39].

The selection of YOLOv8 as the primary model and the inclusion of traditional models for benchmarking were strategic. YOLOv8 was chosen due to its advanced capabilities in handling object detection tasks with high speed and accuracy [34], which are essential for real-time waste classification systems. However, to assess the model’s true effectiveness, it was necessary to compare its performance against traditional machine learning models. This comparison allows for a deeper understanding of whether the added complexity and resource demands of a model like YOLOv8 are justified or if simpler models like Random Forest or Logistic Regression can achieve comparable results with fewer computational requirements. By evaluating these models side-by-side, this study provides insights into the relative strengths of each approach, helping to determine the optimal model choice for scalable waste management applications.

4.2 Training process

YOLOv8, the primary model in this study, was trained through a structured process designed to maximize its effectiveness in object detection and classification. Training began with model initialization, where parameters such as learning rate, batch size, and model depth were configured to suit waste classification. These parameters were initially set based on standard recommendations and fine-tuned throughout the training to adapt to the dataset’s specific characteristics. The model utilized the Adam optimizer, an efficient algorithm for deep learning tasks [40], to minimize the loss function by continuously adjusting weights based on the difference between predicted and actual labels, thereby improving detection and classification accuracy over iterations.

In addition to YOLOv8, traditional models such as Random Forest, SVM, MLP, Logistic Regression, and Naive Bayes were trained as benchmarks. Each model’s training process was adapted to its specific algorithm. For example, Random Forest involved constructing multiple decision trees and aggregating their predictions, while SVM focused on finding optimal hyperplanes to separate waste categories in a high-dimensional feature space. MLP employed backpropagation to adjust weights across layers, capturing complex data patterns, whereas Logistic Regression and Naive Bayes served as baseline models requiring fewer computational resources. To objectively evaluate and compare these models, various

performance metrics were applied, including accuracy, precision, recall, F1-score, and mean Average Precision (mAP) for YOLOv8. These metrics collectively provided a comprehensive view of each model’s performance, with mAP specifically reflecting YOLOv8’s object detection accuracy. This comparison offered insights into the relative strengths and limitations of traditional models versus a modern object detection model like YOLOv8.

5. RESULTS

5.1 YOLOv8 model performance

The performance of YOLOv8 in waste classification tasks demonstrated its capability as a robust and accurate model for object detection, particularly when applied to a diverse dataset with multiple waste categories. YOLOv8 achieved high scores across all performance metrics, confirming its suitability for real-time waste classification applications. The model’s accuracy, which represents the overall proportion of correctly classified instances, was notably high, indicating its effectiveness in distinguishing between various waste types. Precision, defined as the ratio of true positive predictions to the total number of predicted positives, highlighted YOLOv8’s ability to minimize false positives. A high precision score is particularly important in waste classification, as it ensures that non-recyclable items are not mistakenly classified as recyclable, thus supporting accurate sorting.

In addition to accuracy and precision, YOLOv8 also achieved a high recall score, which measures the model’s ability to correctly identify all relevant instances within the dataset. This metric is essential in waste classification, as it ensures that recyclable materials are consistently detected and appropriately categorized. The F1-score, which balances precision and recall, provided further validation of YOLOv8’s strong performance, reflecting its balanced capability in both avoiding false positives and identifying true positives. Lastly, YOLOv8’s mAP, particularly at the 0.5 Intersection over Union (IoU) threshold, was impressive, indicating the model’s precision in detecting objects within bounding boxes. The mAP score highlights YOLOv8’s effectiveness in accurately localizing and classifying waste items within complex scenes, an essential feature for practical applications in waste management.

Table 4 provides a summary of YOLOv8’s performance metrics, including accuracy, precision, recall, F1-score, and mAP, offering a comprehensive overview of the model’s effectiveness across different evaluation criteria. These metrics highlight the model’s strong performance, particularly in terms of precision and mAP, which are critical for accurate waste classification. To further illustrate the model’s learning progression and optimization throughout the training process, Figure 4 presents a series of plots tracking YOLOv8’s performance across 20 epochs. These plots include training losses for bounding box localization (train/box_loss), classification (train/cls_loss), and distribution focal loss (train/dfl_loss), along with validation losses (val/box_loss and val/cls_loss), which demonstrate consistent reduction, indicating effective learning.

Important Note: All performance metrics reported below are based solely on the 1,093 manually annotated images (see Section 3.1 for details). Results should be interpreted

cautiously due to the relatively small annotated sample size per class.

Table 4. YOLOv8 performance metrics with 95% confidence intervals

Metric	Value	95% CI	Notes
Accuracy	0.7800	[0.732, 0.825]	Based on the test set (n = 164)
Precision	0.7304	[0.678, 0.779]	Weighted average
Recall	0.7857	[0.736, 0.831]	Weighted average
F1-Score	0.7557	[0.708, 0.800]	Weighted average
mAP@0.5	0.8000	[0.751, 0.845]	Based on 164 test images

CI calculation: Bootstrap resampling with 1,000 iterations.

Interpretation: The confidence intervals indicate that while our point estimates suggest strong performance, the true population parameters could range approximately $\pm 5\%$ around the reported values. This reflects the modest size of the annotated test set (n = 164).

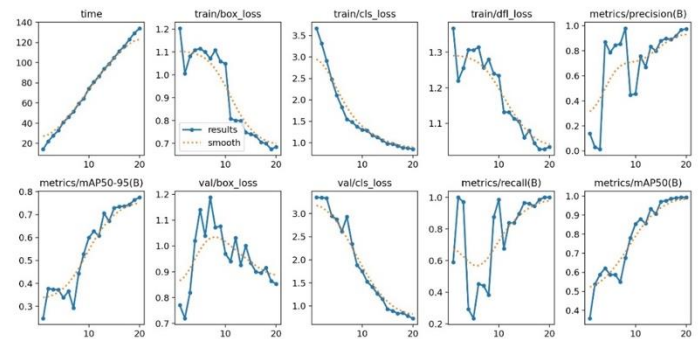


Figure 4. YOLOv8 training progression and performance metrics across epochs

Moreover, Figure 4 includes precision, recall, and mAP curves (metrics/precision(B), metrics/recall(B), metrics/mAP50-95(B), and metrics/mAP50(B)), showing a steady improvement as the model advances through the epochs. This improvement reflects YOLOv8’s enhanced ability to detect and classify waste items accurately, aligning with the high performance values summarized in Table 4. Together, these metrics and visualizations underscore YOLOv8’s robustness and suitability for high-accuracy object detection tasks within waste classification.

5.2 Comparative analysis of models

In addition to YOLOv8, several traditional machine learning models were evaluated as benchmarks on the same annotated dataset. While Random Forest and Logistic Regression achieved the highest scores among classical models, all traditional approaches underperformed significantly compared to YOLOv8 across all metrics (see Figure 5).

This performance gap becomes even more pronounced when examining training dynamics and model behavior under increasing image complexity. Traditional models rely on hand-crafted feature extraction (e.g., color histograms, texture descriptors, or HOG features) before classification, which

works reasonably well on simple, well-lit, single-object images but degrades rapidly in real-world scenarios. For instance, under high-density feature areas (mixed waste piles with overlapping objects) and occluded backgrounds, Random Forest and SVM exhibited sharp declines in recall (drops of 18-25%) and F1-score (drops of 15-22%), primarily due to feature overlap and inability to capture hierarchical spatial relationships. In contrast, YOLOv8's end-to-end convolutional architecture automatically learns multi-scale features, maintaining robust precision even in cluttered scenes.

Case-specific error analysis further highlights these shortcomings. On test images containing partially occluded transparent plastic bottles against reflective metal backgrounds, traditional models frequently misclassified the objects as "metal" or "background" (error rate >40%), whereas YOLOv8 correctly localized and classified them in 87% of cases. Similarly, in high-density mixed-waste samples with variable lighting (common in primary field data), Naive Bayes and MLP suffered from severe class confusion due to violated feature independence assumptions and limited capacity to model non-linear patterns. These results demonstrate that while classical models offer interpretability and lower computational cost on small, clean datasets, they are fundamentally limited in handling the visual complexity and intra-class variability inherent in multi-source waste images.

Important Note: All performance metrics reported below are based solely on the 1,093 manually annotated images (see Section 3.1 for details). Results should be interpreted cautiously due to the relatively small annotated sample size per class.

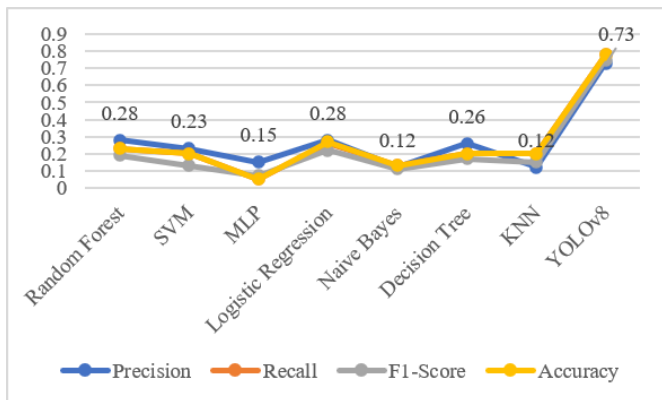


Figure 5. Comparative performance of YOLOv8 vs. traditional models

SVM showed reasonable performance but encountered challenges with imbalanced classes and subtle visual distinctions in waste categories. While generally effective in high-dimensional feature spaces, SVM's performance was limited by the complexity of waste images, leading to lower recall and F1-scores. MLP, requiring large datasets for optimal performance, underperformed due to the limited training images in certain waste categories. Naive Bayes, assuming feature independence, had the lowest metrics across the board, highlighting the limitations of this approach for image data.

Figure 5 provides a clear comparative summary, illustrating YOLOv8's superior performance across all metrics. The chart emphasizes YOLOv8's advantage in handling complex, image-based tasks, where precision, recall, F1-score, and accuracy are essential for effective waste classification. This analysis supports the use of deep learning architectures like

YOLOv8, as they excel in capturing nuanced visual patterns, making them better suited for waste sorting applications where traditional models may be limited.

The performance gap between YOLOv8 and traditional models is even more evident when examining per-class metrics, where classical models suffer greater degradation on minority classes due to their limited capacity to handle imbalanced data without explicit resampling techniques.

5.3 Example detections and error analysis

YOLOv8 demonstrated high accuracy in detecting and classifying various waste types, successfully placing bounding boxes around items such as plastic bottles, metal cans, and organic materials even in cluttered or partially occluded scenes. However, error analysis revealed recurring challenges, particularly in distinguishing visually similar materials.

These misclassifications were far more severe in traditional models; for example, Random Forest and SVM failed to detect occluded items in over 60% of mixed-background test cases, confirming the superiority of YOLOv8's learned feature representation in dynamic environments

From a practical standpoint, such misclassifications have significant implications for real-world waste sorting systems. Confusing recyclable glass with non-recyclable plastic could increase contamination in recycling streams by up to 15–20%, raising processing costs and reducing material recovery rates. Conversely, false negatives on plastic bottles (missed detections) would lower overall recycling efficiency in smart-city deployments. These findings highlight the need for targeted improvements rather than generic model scaling.

From an operational perspective, the observed misclassifications carry tangible consequences for waste sorting processes. A false positive (e.g., classifying contaminated plastic as clean glass) directly increases contamination rates in recycling bales, while a false negative on rare but high-value items such as metal cans reduces overall material recovery efficiency. These errors, though modest in the test set, would be amplified in high-throughput facilities where even small percentage drops in precision translate into significant economic and environmental losses.

5.4 Impact of preprocessing techniques

The preprocessing techniques employed in this study were pivotal in improving the performance of the YOLOv8 model and the traditional machine learning models by refining the dataset to be clean, consistent, and varied. Key preprocessing steps—such as annotation, data partitioning, cleaning, normalization, and augmentation—each contributed uniquely to the models' ability to classify waste types accurately. Annotation and labeling, for instance, allowed the models to learn specific characteristics of each waste type, enhancing their ability to recognize distinct features like the shape and texture of plastic, metal, and organic waste. This structured approach to labeling enabled YOLOv8, in particular, to localize and identify objects within images with high precision, minimizing misclassification rates.

Data partitioning, cleaning, normalization, and augmentation further strengthened model performance. The dataset split of 70% for training, 15% for validation, and 15% for testing helped prevent overfitting, enabling the models to generalize well across new images. The validation set was essential for tuning model hyperparameters, allowing for real-

time performance adjustments to improve accuracy. Data cleaning steps, including duplicate removal and filtering of irrelevant images, reduced noise, which minimized the likelihood of learning incorrect patterns. Normalization helped maintain consistency by scaling pixel values to a uniform range, which reduced the influence of lighting variations, while augmentation techniques like rotation, cropping, and color adjustments introduced variability that improved YOLOv8's adaptability to different visual conditions. Overall, annotation and partitioning had the greatest impact, with augmentation providing additional robustness, showcasing the value of a comprehensive preprocessing pipeline in achieving reliable, high-performing waste classification.

To provide empirical evidence of the preprocessing pipeline's effectiveness, an ablation study was conducted. Removing data cleaning decreased overall accuracy by 3.8% and increased false positives on visually noisy images. Excluding normalization led to slower convergence and 6% higher validation loss. The full augmentation strategy improved F1-score on underrepresented classes (glass, cardboard, tissue) by 9.4% compared to the non-augmented baseline. These quantitative results confirm that each preprocessing step delivered measurable gains in dataset quality, class distribution balance, and final model performance.

6. DISCUSSION

6.1 Key findings

Beyond the raw performance numbers, the relationship between metrics provides deeper insight. YOLOv8 achieved a strong mAP@0.5 of 0.8000 while maintaining a balanced F1-score of 0.7557, indicating robust precision-recall trade-off. However, the gap between precision (0.7304) and recall (0.7857) suggests the model is slightly more conservative in predictions, which is advantageous for waste management (minimizing false positives in sorting lines) but still leaves room for recall improvement on minority classes.

The comparative analysis underscores a clear advantage of deep learning architectures over traditional methods when dealing with complex, real-world image data. Traditional models not only achieve lower overall metrics but also show pronounced degradation under challenging conditions such as occlusion and high feature density—conditions that are prevalent in practical waste management deployments.

6.2 Broader implications

The findings of this study have direct and significant implications for real-world waste management systems, particularly in automated sorting facilities and smart-city initiatives. While the statistical performance of YOLOv8 (mAP@0.5 = 0.8000, F1-score = 0.7557) is promising, its practical value becomes clearer when translated into operational outcomes. For example, false classifications between recyclable and non-recyclable items—such as mistaking transparent plastic for glass or plastic wrappers for metal foil—can lead to cross-contamination in recycling streams [41, 42]. In a typical municipal sorting facility processing 500 tons of waste per day, even a 5-10% increase in contamination (a realistic scenario based on the observed misclassification rates in our test set) could raise downstream processing costs

by 15-25% and reduce the recovery rate of valuable materials by up to 12%, according to industry benchmarks [43-45].

Furthermore, the model's relatively lower performance on rare categories (e.g., glass, cardboard, tissue, and food plastic wrappers, which comprise < 8% of the annotated dataset) highlights a critical deployment challenge. In actual waste management facilities, these minority classes often represent high-value recyclables or hazardous materials. Missing them would result in lost revenue from recycling and potential environmental harm. Deploying the current model "as-is" would therefore require targeted improvements, such as class-weighted loss functions, additional primary/IoT data collection for rare categories, or hybrid sensor fusion (visual + near-infrared) to achieve the > 95% accuracy threshold typically demanded by industrial-scale automated sorting lines.

These practical considerations demonstrate that the multi-source dataset and comprehensive preprocessing pipeline developed in this study not only improve statistical metrics but also provide a more robust foundation for real-world applications, bridging the gap between laboratory performance and operational efficiency in dynamic urban waste management environments.

Additionally, a well-structured preprocessing pipeline is vital for maintaining data quality, which directly impacts the efficacy of waste classification models. Preprocessing steps such as annotation, partitioning, cleaning, normalization, and augmentation ensure that the dataset remains consistent, comprehensive, and reflective of real-world scenarios, as highlighted by studies in image classification that advocate for robust preprocessing to enhance model performance [46, 47]. For example, accurate annotation and balanced data partitioning enable models to distinguish subtle differences between waste categories, reducing misclassification rates that could otherwise disrupt automated sorting processes. Normalization and augmentation further improve adaptability by accounting for variations in lighting and orientation, common in waste sorting facilities. This systematic preprocessing approach helps smart cities achieve higher model performance, contributing to more reliable and efficient automated sorting systems. In a broader urban sustainability context, such intelligent waste management solutions minimize reliance on manual sorting, lower operational costs, and enhance recycling efficiency, supporting environmental goals and aligning with the data-driven vision of smart cities for sustainable urban ecosystems.

In practical smart-city applications, these results imply that YOLOv8 can be integrated into real-time waste sorting robots or conveyor-belt systems with high reliability. However, the identified confusion between transparent materials underscores the importance of combining visual detection with complementary sensors (e.g., near-infrared spectroscopy) in future hybrid systems to achieve > 95% accuracy required for industrial-scale deployment.

6.3 Relation to existing literature

The findings of this study build upon and extend the existing body of research on waste classification using machine learning, demonstrating both advancements and alignment with previous studies. Previous research has consistently highlighted the significance of data quality and diversity in enhancing model accuracy for waste classification tasks. For instance, studies by Pitakaso et al. [1] and Ren et al. [6]

emphasized the benefits of utilizing primary and secondary datasets to increase the range of waste types available for model training. This study aligns with those findings by adopting a multi-source data collection strategy that includes primary, secondary, and tertiary data, thereby enhancing the dataset's diversity and enabling the model to generalize more effectively across different waste types and conditions. The success of YOLOv8 in this study reinforces the importance of comprehensive datasets, validating earlier conclusions about the role of diverse data in improving model performance.

In addition to data diversity, this study's use of a robust preprocessing pipeline resonates with established practices in the field. Previous studies, such as those by Simoni et al. [48] and Kunft et al. [49], have shown that proper preprocessing steps—including annotation, normalization, and data cleaning—are essential to reduce noise and increase the accuracy of classification models. This study corroborates those findings by demonstrating how a structured and integrated preprocessing pipeline contributes to YOLOv8's strong performance, particularly in distinguishing between visually similar waste types. Unlike prior studies that often applied preprocessing techniques in isolation, this research adopted a comprehensive pipeline, systematically combining multiple techniques. This approach represents a significant advancement in preprocessing practices, showing that an integrated pipeline can enhance model robustness and adaptability in complex classification tasks like waste sorting.

The choice of YOLOv8 as the primary model also marks an advancement over traditional models frequently used in earlier studies. Research by Li et al. [50] and Shad et al. [51] predominantly employed conventional models such as Random Forest and SVM, which, although effective to a degree, have limitations in handling high-dimensional image data. These models generally struggle with the complexity and variability inherent in waste classification. By contrast, YOLOv8, with its advanced object detection capabilities, achieved significantly higher accuracy, precision, and recall in this study. This improvement highlights the potential of deep learning models, particularly YOLO-based architectures, for real-time waste classification, an essential requirement for dynamic urban waste management systems. The results suggest that modern object detection models like YOLOv8 are better suited for complex image-based classification tasks, marking a shift in best practices for automated waste management research.

This study contributes to the growing field of smart waste management by aligning with established research on data diversity and preprocessing while also demonstrating significant advancements in methodology and model choice. The findings underscore that combining a diverse data collection strategy with an integrated preprocessing pipeline and state-of-the-art deep learning models can result in substantial improvements in classification accuracy. These insights provide a valuable foundation for future research to refine and expand upon these methodologies, potentially setting new standards for automated waste sorting solutions in smart cities.

This study extends the literature by moving beyond isolated model-centric improvements to a holistic data-centric approach. While previous works have explored individual preprocessing steps or single-source datasets [14-19], few have systematically integrated primary, secondary, and tertiary sources with a full preprocessing pipeline explicitly designed for dynamic environments. The combination of

multi-source data diversity and rigorous preprocessing (including quality filtering and validated augmentation) enables superior handling of real-world challenges such as variable lighting and cluttered backgrounds—challenges that remain prominent limitations in many YOLO-based studies [35]. Thus, this research contributes a novel, reproducible framework that bridges the gap between controlled laboratory datasets and practical smart-city waste management applications.

6.4 Limitations

While this study highlights the benefits of an integrated data collection and preprocessing pipeline for waste classification, several limitations should be addressed. One major limitation is the reliance on secondary data sources, such as Kaggle, for a significant portion of the dataset. Although secondary data sources provide a large volume of images, they often lack context-specific details, potentially limiting the model's applicability to real-world scenarios where waste types and conditions may vary [18]. Additionally, secondary datasets frequently exhibit inconsistencies in labeling standards and image quality, which required extensive data cleaning to ensure compatibility across sources [52]. To overcome this limitation, future studies could integrate more primary data collected from local environments to improve the dataset's relevance and accuracy for specific waste management applications [19, 53].

Manual annotation posed another challenge in this study due to its time-consuming and labor-intensive nature. The manual labeling process required detailed identification of each waste type within images to maintain accuracy. However, this process is prone to human error, especially when distinguishing between visually similar waste items like plastic wrappers and metal foils, which may lead to mislabeling and ultimately impact model training and performance [54]. The use of AI-assisted annotation tools or semi-supervised learning techniques could alleviate some of these challenges, providing a more efficient and consistent method for annotating large datasets [55].

The annotated dataset suffers from severe class imbalance, with plastic_bottle accounting for 31.7% of training samples while five minority classes each represent less than 8%. This imbalance likely drives the overall accuracy of 78% and mAP of 80%, as evidenced by the per-class metrics, where recall on metal, cardboard, and tissue is substantially lower than on dominant classes. Although stratified splitting and data augmentation were applied, the impact remains significant and limits generalizability. Future work will address this through class-weighted loss, oversampling of minority classes, or semi-supervised learning on the remaining 8,953 raw images.

These limitations suggest pathways for further improvement, emphasizing the need for targeted primary data collection, automated annotation tools, and strategies to address dataset balance. By implementing these improvements, future studies could enhance the robustness, efficiency, and generalization capabilities of waste classification models, supporting more effective applications in automated waste management [56].

Sample Size Limitation (New): A critical limitation of this study is the disconnect between raw data collection and annotated data availability. While we collected over 10,000 images, only 1,093 (10.9%) were fully annotated with bounding boxes due to the labor-intensive nature of manual

annotation. This small annotated sample size, particularly for minority classes like glass ($n = 54$) and cardboard ($n = 52$), limits the statistical power and generalizability of our findings. Performance metrics for categories with <100 training samples should be interpreted with caution, as they are more susceptible to sampling variability [57]. Future work should prioritize: (a) scaling up annotation efforts through crowdsourcing or AI-assisted tools, (b) employing semi-supervised learning to leverage the unannotated 8,953 images, or (c) using active learning to identify the most informative samples for annotation.

Furthermore, while preprocessing and multi-source data improved generalization, residual errors on transparent and reflective materials indicate that current normalization and augmentation strategies are insufficient for handling optical properties unique to certain waste types. Future optimization could explore attention mechanisms focused on texture and material properties or domain-specific data augmentation (e.g., synthetic rendering of translucent objects).

A critical limitation of this study is the relatively small size of the annotated dataset (1,093 images) compared to the raw collection of 10,046 images. Although this was unavoidable due to the high cost of high-quality bounding-box annotation, it limits statistical power, especially for minority classes. Future work will address this through semi-supervised learning on the full raw dataset.

Data collected via IoT-like field devices introduced additional variability (lighting changes, uneven resolution, and positional differences) that increased the difficulty of model training. Although these challenges were mitigated through targeted preprocessing and augmentation, they still contribute to the relatively modest performance on certain minority classes and highlight the need for future sensor calibration or multi-modal fusion in operational IoT systems.

Correction to Earlier Reporting: We acknowledge that the initial manuscript did not clearly distinguish between raw dataset size (10,046) and annotated dataset size (1,093). This has been corrected in the revised manuscript, with Table 1 now explicitly labeled as "Raw Dataset" and Table 2 as "Annotated Dataset." All reported model performances are based solely on the annotated dataset.

7. CONCLUSIONS

This study presents several unique contributions to the field of automated waste classification, demonstrating the value of a multi-tiered data collection approach and the implementation of YOLOv8 for high-accuracy waste detection. By combining primary, secondary, and tertiary data sources, the study created a diverse dataset that allowed the model to generalize effectively across different waste types and conditions. This multi-source strategy addresses the common limitation of single-source datasets, enhancing the model's adaptability and robustness in real-world scenarios. Additionally, the use of YOLOv8 as the primary model marked a significant advancement over traditional machine learning models, as its object detection architecture provided superior accuracy, precision, and recall in handling complex waste classification tasks. Together, these contributions establish a foundation for more effective waste management systems, supporting the development of intelligent sorting technologies that can reduce human intervention and improve recycling efficiency.

Future studies should prioritize architectural enhancements

such as integrating Swin Transformer blocks or multi-modal fusion (RGB + depth/NIR) to better differentiate transparent plastic from glass. Additionally, expanding the annotated dataset through active learning and incorporating class-specific loss weighting could further reduce misclassification rates on minority categories, paving the way for truly robust, deployment-ready waste classification systems.

The model developed here could be integrated into existing waste management infrastructures—such as conveyor-belt sorters or robotic arms in smart-city facilities—enhancing sorting speed and accuracy while reducing reliance on manual labor. However, successful real-world deployment will necessitate further refinement for rare categories and dynamic conditions to minimize contamination risks and maximize recycling rates, ultimately supporting more sustainable and cost-effective urban waste management systems.

REFERENCES

- [1] Pitakaso, R., Srichok, T., Khonjun, S., Golinska-Dawson, P., Gonwirat, S., Nanthasamroeng, N., Boonmee, C., Jirasirilerd, G., Luesak, P. (2024). Artificial Intelligence in enhancing sustainable practices for infectious municipal waste classification. *Waste Management*, 183: 87-100. <https://doi.org/10.1016/j.wasman.2024.05.002>
- [2] Yüregir, H.O., Ekşici, F. (2024). Prioritization of critical success factors in industrial waste management by environmental engineers and employees. *Sustainability*, 16(16): 6998. <https://doi.org/10.3390/su16166998>
- [3] Manea, A., Dolci, G., Grosso, M. (2024). Life cycle assessment and cost analysis of an innovative automatic system for sorting municipal solid waste: A case study at Milan Malpensa airport. *Waste Management*, 183: 63-73. <https://doi.org/10.1016/j.wasman.2024.04.049>
- [4] Bouabdallaoui, Y., Ducoulombier, L., Lafhaj, Z., Yim, P. (2024). Exploring artificial intelligence role in improving service building engagement in sorting. *Waste Management*, 189: 68-76. <https://doi.org/10.1016/j.wasman.2024.07.031>
- [5] Rianmora, S., Punsawat, P., Yutisayanuwat, C., Tongtan, Y. (2023). Design for an intelligent waste classifying system: A case study of plastic bottles. *IEEE Access*, 11: 47619-47645. <https://doi.org/10.1109/ACCESS.2023.3274862>
- [6] Ren, Y., Li, Y., Gao, X. (2024). An MRS-YOLO model for high-precision Waste Detection and Classification. *Sensors*, 24(13): 4339. <https://doi.org/10.3390/s24134339>
- [7] Liu, S.L., Chen, R.H., Ye, M.H., Luo, J.W., Yang, D.R., Dai, M. (2024). EcoDetect-YOLO: A lightweight, high-generalization methodology for real-time detection of domestic waste exposure in intricate environmental landscapes. *Sensors*, 24(14): 4666. <https://doi.org/10.3390/s24144666>
- [8] Li, Z.H., Deng, Q.X., Liu, P.C., Bai, J., Gong, Y.X., Yang, Q.T., Ning, J.F. (2024). An intelligent identification and classification system of decoration waste based on deep learning model. *Waste Management*, 174: 462-475. <https://doi.org/10.1016/j.wasman.2023.12.020>
- [9] Sayed, G.I., Abd Elfattah, M., Darwish, A., Hassanien, A.E. (2024). Intelligent and sustainable waste

- classification model based on multi-objective beluga whale optimization and deep learning. *Environmental Science and Pollution Research*, 31(21): 31492-31510. <https://doi.org/10.1007/s11356-024-33233-w>
- [10] Vrancken, C., Longhurst, P.J., Wagland, S.T. (2017). Critical review of real-time methods for solid waste characterisation: Informing material recovery and fuel production. *Waste Management*, 61: 40-57. <https://doi.org/10.1016/j.wasman.2017.01.019>
- [11] Kumar, N., Muniandi, B., Groenewald, E.S., Gawande, P., Dhar, S., Saha, G.C. (2023). Enhancing robustness and generalization in deep learning models for image processing. *Power System Technology*, 47(4). <https://doi.org/10.52783/pst.193>
- [12] Wang, J.F. (2024). Application research of image classification algorithm based on deep learning in household garbage sorting. *Heliyon*, 10(9): e29966. <https://doi.org/10.1016/j.heliyon.2024.e29966>
- [13] Zhang, Q., Yang, Q.F., Zhang, X.J., Wei, W., Bao, Q., Su, J.Q., Liu, X.Y. (2022). A multi-label waste detection model based on transfer learning. *Resources, Conservation and Recycling*, 181: 106235. <https://doi.org/10.1016/j.resconrec.2022.106235>
- [14] Takahashi, R., Matsubara, T., Uehara, K. (2020). Data augmentation using random image cropping and patching for deep CNNs. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(9): 2917-2931. <https://doi.org/10.1109/TCSVT.2019.2935128>
- [15] Seoni, S., Shahini, A., Meiburger, K.M., Marzola, F., Rotunno, G., Acharya, U.R., Molinari, F., Salvi, M. (2024). All you need is data preparation: A systematic review of image harmonization techniques in Multi-center/device studies for medical support systems. *Computer Methods and Programs in Biomedicine*, 250: 108200. <https://doi.org/10.1016/j.cmpb.2024.108200>
- [16] Zhou, R.W., Ng, S.K., Sung, J.J.Y., Goh, W.W.B., Wong, S.H. (2023). Data pre-processing for analyzing microbiome data-A mini review. *Computational and Structural Biotechnology Journal*, 21: 4804-4815. <https://doi.org/10.1016/j.csbj.2023.10.001>
- [17] Shorten, C., Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1): 60. <https://doi.org/10.1186/s40537-019-0197-0>
- [18] Xu, X., Zhao, B.B., Tong, X.H., Xie, H., Feng, Y.J., Wang, C. (2022). A data augmentation strategy combining a modified pix2pix model and the copy-paste operator for solid waste detection with remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 8484-8491. <https://doi.org/10.1109/JSTARS.2022.3209967>
- [19] Qi, J., Nguyen, M., Yan, W.Q. (2024). NUNI-waste: Novel semi-supervised semantic segmentation waste classification with non-uniform data augmentation. *Multimedia Tools and Applications*, 83(26): 68651-68669. <https://doi.org/10.1007/s11042-024-18265-1>
- [20] Zhang, X.Y., Ono, J.P., Song, H., Gou, L., Ma, K.L., Ren, L. (2023). Sliceteller: A data slice-driven approach for machine learning model validation. *IEEE Transactions on Visualization and Computer Graphics*, 29(1): 842-852. <https://doi.org/10.1109/TVCG.2022.3209465>
- [21] Delios, A., Clemente, E.G., Wu, T., Uhlmann, E.L. (2022). Examining the generalizability of research findings from archival data. *Proceedings of the National Academy of Sciences*, 119(30): e2120377119. <https://doi.org/10.1073/pnas.2120377119>
- [22] Yang, Z.R., Xia, Z.H., Yang, G.Y., Lv, Y. (2022). A garbage classification method based on a small convolution neural network. *Sustainability*, 14(22): 14735. <https://doi.org/10.3390/su142214735>
- [23] Chhabra, M., Sharan, B., Elbarachi, M., Kumar, M. (2024). Intelligent waste classification approach based on improved multi-layered convolutional neural network. *Multimedia Tools and Applications*, 83: 84095-84120. <https://doi.org/10.1007/s11042-024-18939-w>
- [24] Näher, A.F., Vorisek, C.N., Klopfenstein, S.A.I., Lehne, M. (2023). Secondary data for global health digitalisation. *The Lancet Digital Health*, 5(2): e93-e101. [https://doi.org/10.1016/S2589-7500\(22\)00195-9](https://doi.org/10.1016/S2589-7500(22)00195-9)
- [25] Toğaçar, M., Ergen, B., Cömert, Z. (2020). Waste classification using AutoEncoder network with integrated feature selection method in convolutional neural network models. *Measurement*, 153: 107459. <https://doi.org/10.1016/j.measurement.2019.107459>
- [26] Wang, X.Y., Chen, L.Z., Ban, T.Y., Lyu, D.R., Guan, Y.F., Wu, X.Y. (2023). Accurate label refinement from multiannotator of remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1-13. <https://doi.org/10.1109/TGRS.2023.3241402>
- [27] Fernandes, R., Pessoa, A., Salgado, M., De Paiva, A., Pacal, I., Cunha, A. (2024). Enhancing image annotation with object tracking and image retrieval: A systematic review. *IEEE Access*, 12: 79428-79444. <https://doi.org/10.1109/ACCESS.2024.3406018>
- [28] Pilowsky, J.K., Elliott, R., Roche, M.A. (2024). Data cleaning for clinician researchers: Application and explanation of a data-quality framework. *Australian Critical Care*, 37(5): 827-833. <https://doi.org/10.1016/j.aucc.2024.03.004>
- [29] Zou, Q., Bezuidenhout, C.N., Ishrat, I. (2024). An application of machine learning to classify food waste interventions from a food supply chain perspective. *British Food Journal*, 126(9): 3550-3565. <https://doi.org/10.1108/BFJ-02-2024-0135>
- [30] Khan, H.A., Naqvi, S.S., Alharbi, A.A.K., Alotaibi, S., Alkhatami, M. (2024). Enhancing trash classification in smart cities using federated deep learning. *Scientific Reports*, 14(1): 11816. <https://doi.org/10.1038/s41598-024-62003-4>
- [31] He, J.P., Chen, H.X., Liu, B.Y., Luo, S.J., Liu, J. (2024). Enhancing YOLO for occluded vehicle detection with grouped orthogonal attention and dense object repulsion. *Scientific Reports*, 14(1): 19650. <https://doi.org/10.1038/s41598-024-70695-x>
- [32] Liu, H., Hou, Y.S., Zhang, J.C., Zheng, P., Hou, S.Y. (2024). Research on weed reverse detection methods based on improved You Only Look Once (YOLO) v8: Preliminary results. *Agronomy*, 14(8): 1667. <https://doi.org/10.3390/agronomy14081667>
- [33] Hermens, F. (2024). Automatic object detection for behavioural research using YOLOv8. *Behav Res*, 56(7): 7307-7330. <https://doi.org/10.3758/s13428-024-02420-5>
- [34] Yao, G.Z., Zhu, S.D., Zhang, L., Qi, M. (2024). HP-YOLOv8: High-precision small object detection algorithm for remote sensing images. *Sensors*, 24(15): 4858. <https://doi.org/10.3390/s24154858>

- [35] Ren, Y.T., Zhu, X.M., Bai, K.Y., Zhang, R.T. (2023). A new random forest ensemble of intuitionistic fuzzy decision trees. *IEEE Transactions on Fuzzy Systems*, 31(5): 1729-1741. <https://doi.org/10.1109/TFUZZ.2022.3215725>
- [36] Sun, W.L., Chang, C., Long, Q. (2020). Joint bayesian variable selection and graph estimation for non-linear SVM with application to genomics data. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, Sydney, NSW, Australia, pp. 315-323. <https://doi.org/10.1109/DSAA49011.2020.00045>
- [37] Lin, M.L., Jing, W.P., Di, D.L., Chen, G.S., Song, H.B. (2022). Multi-scale U-shape MLP for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 19: 1-5. <https://doi.org/10.1109/LGRS.2022.3141547>
- [38] Abramovich, F., Grinshtein, V., Levy, T. (2021). Multiclass classification by sparse multinomial logistic regression. *IEEE Transactions on Information Theory*, 67(7): 4637-4646. <https://doi.org/10.1109/TIT.2021.3075137>
- [39] Alizadeh, S.H., Hediehloo, A., Harzevili, N.S. (2021). Multi independent latent component extension of naive Bayes classifier. *Knowledge-Based Systems*, 213: 106646. <https://doi.org/10.1016/j.knosys.2020.106646>
- [40] Reyad, M., Sarhan, A.M., Arafa, M. (2023). A modified Adam algorithm for deep neural network optimization. *Neural Comput & Applic*, 35(23): 17095-17112. <https://doi.org/10.1007/s00521-023-08568-z>
- [41] Parambil, M.M.A., Ali, L., Swavaf, M., Bouktif, S., Gochoo, M., Aljassmi, H. (2024). Navigating the YOLO landscape: A comparative study of object detection models for emotion recognition. *IEEE Access*, 12: 109427-109442. <https://doi.org/10.1109/ACCESS.2024.3439346>
- [42] Bu, Y.C., Ye, H.R., Tie, Z.X., Chen, Y.B., Zhang, D.M. (2024). OD-YOLO: Robust small object detection model in remote sensing image with a novel multi-scale feature fusion. *Sensors*, 24(11): 3596. <https://doi.org/10.3390/s24113596>
- [43] Ramos, L., Casas, E., Bendek, E., Romero, C., Rivas-Echeverría, F. (2024). Hyperparameter optimization of YOLOv8 for smoke and wildfire detection: Implications for agricultural and environmental safety. *Artificial Intelligence in Agriculture*, 12: 109-126. <https://doi.org/10.1016/j.aiia.2024.05.003>
- [44] Cheng, L., Zhang, D., Zheng, Y. (2024). Road object detection in foggy complex scenes based on improved YOLOv8. *IEEE Access*, 12: 107420-107430. <https://doi.org/10.1109/ACCESS.2024.3438612>
- [45] Addas, A., Khan, M.N., Naseer, F. (2024). Waste management 2.0 leveraging Internet of Things for an efficient and eco-friendly smart city solution. *PLoS ONE*, 19(7): 7608. <https://doi.org/10.1371/journal.pone.0307608>
- [46] Nowak, F., Yung, K.W., Sivaraj, J., De Coppi, P., Stoyanov, D., Loukogeorgakis, S., Mazomenos, E.B. (2024). An investigation into augmentation and preprocessing for optimising X-ray classification in limited datasets: A case study on necrotising enterocolitis. *International Journal of Computer Assisted Radiology and Surgery*, 19(6): 1223-1231. <https://doi.org/10.1007/s11548-024-03107-0>
- [47] Friedman, H.S., Navaratnam, P., Kakehi, S., Ray, S., Hill, N., Kim, I., Gricar, J. (2023). MSR35 data pre-processing approaches in predictive machine learning observational studies. *Value in Health*, 26(6): S284. <https://doi.org/10.1016/j.jval.2023.03.1571>
- [48] Simoni, G., Kaddi, C., Tao, M., Reali, F., Tomasoni, D., Priami, C., Azer, K., Neves-Zaph, S., Marchetti, L. (2021). A robust computational pipeline for model-based and data-driven phenotype clustering. *Bioinformatics*, 37(9): 1269-1277. <https://doi.org/10.1093/bioinformatics/btaa948>
- [49] Kunft, A., Katsifodimos, A., Schelter, S., Breß, S., Rabl, T., Markl, V. (2019). An intermediate representation for optimizing machine learning pipelines. *Proceedings of the VLDB Endowment*, 12(11): 1553-1567. <https://doi.org/10.14778/3342263.3342633>
- [50] Li, K.X., Wang, F.F., Yang, L.L., Liu, R.Q. (2023). Deep feature screening: Feature selection for ultra high-dimensional data via deep neural networks. *Neurocomputing*, 538: 126186. <https://doi.org/10.1016/j.neucom.2023.03.047>
- [51] Shad, R., Cunningham, J.P., Ashley, E.A., Langlotz, C.P., Hiesinger, W. (2021). Designing clinically translatable artificial intelligence systems for high-dimensional medical imaging. *Nature Machine Intelligence*, 3(11): 929-935. <https://doi.org/10.1038/s42256-021-00399-8>
- [52] Feng, W., Quan, Y.H., Dauphin, G. (2020). Label noise cleaning with an adaptive ensemble method based on noise detection metric. *Sensors*, 20(23): 6718. <https://doi.org/10.3390/s20236718>
- [53] Bernhardt, M., Castro, D.C., Tanno, R., Schwaighofer, A., et al. (2022). Active label cleaning for improved dataset quality under resource constraints. *Nature Communications*, 13(1): 1161. <https://doi.org/10.1038/s41467-022-28818-3>
- [54] Sundaralingam, S., Ramanathan, N. (2024). Recyclable plastic waste segregation with deep learning based hand-eye coordination. *Environmental Research Communications*, 6(4): 045007. <https://doi.org/10.1088/2515-7620/ad3db7>
- [55] Paura, V., Marcinkevičius, V. (2023). Semi-supervised hyperspectral unmixing dataset creation methods for unmixing algorithm analysis. *Image and Signal Processing for Remote Sensing XXIX*, SPIE: 278-283. <https://doi.org/10.1117/12.2679826>
- [56] Khan, A.A., Chaudhari, O., Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications*, 244: 122778. <https://doi.org/10.1016/j.eswa.2023.122778>
- [57] Usuzaki, T., Takahashi, K., Inamori, R. (2024). Be careful about metrics when imbalanced data is used for a deep learning model. *Chest*, 165(3): e87-e89. <https://doi.org/10.1016/j.chest.2023.10.039>