Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI

(Rekayasa Sistem dan Teknologi Informasi)

Vol. 10 No. 3 (2026) 580 - 589 e-ISSN: 2580-0760

Impact of Dataset Background on Deep Learning-Based Waste Classification

Nazzua Azzahra[1], Aditiya Hermawan[2], Junaedi[3*], Yusuf Kurnia[4], Edy[5]
[1,2,4]Informatics Engineering, Faculty of Science and Technology, Buddhi Dharma University, Tangerang, Indonesia
[3]Information Systems, Faculty of Science and Technology, Buddhi Dharma University, Tangerang, Indonesia
[5]Software Engineering, Faculty of Science and Technology, Buddhi Dharma University, Tangerang, Indonesia

*junaedi@ubd.ac.id

Abstract

Accurate waste classification plays a vital role in supporting effective waste management and promoting environmental sustainability, especially amid the continuing increase in global waste generation. This study investigates how the presence and removal of image backgrounds influence the performance of deep learning models in automated waste classification. Two Convolutional Neural Network architectures, namely MobileNetV2 and DenseNet169, were evaluated using a dataset comprising 5,054 images across six waste categories: cardboard, glass, metal, paper, plastic, and trash. Each architecture was trained and tested on two dataset variants: original images with backgrounds and images with the backgrounds removed. Model performance was assessed using accuracy, precision, recall, F1-score, and ROC AUC. The results indicate that DenseNet169 consistently outperformed MobileNetV2 across all evaluation metrics. The highest accuracy, reaching 88.33%, was achieved by DenseNet169 when trained on images retaining their original backgrounds. This suggests that background information may provide meaningful contextual features that enhance classification performance. Conversely, removing backgrounds can limit the visual information available to the model and potentially reduce predictive effectiveness. These findings emphasize the importance of carefully considering background characteristics during dataset preparation and model training. Moreover, the study demonstrates that selecting an appropriate model architecture in relation to dataset properties is essential for optimizing classification outcomes. Overall, this research offers practical insights for improving dataset design and model selection in future automated waste classification systems, while contributing to the advancement of scalable and intelligent deep learning-based waste management solutions.

Keywords: convolutional neural network; deep learning; densenet169; image classification; mobilenetv2

How to Cite: N. Azzahra, A. Hermawan, Junaedi, Y. Kurnia, and Edy, "Impact of Dataset Background on Deep Learning-Based Waste Classification", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 10, no. 3, pp. 580 - 589, Jun. 2026.
*Permalink/*DOI: https://doi.org/10.29207/resti.v10i3.6965

Received: July 15, 2025
Accepted: February 3, 2026
Available Online: June 6, 2026

1. Introduction

Waste management is a serious problem, as population growth and human activities continue to increase. According to the latest report from International Lianhe Zaobao, worldwide waste volume is predicted to increase by as much as 70% by 2050, which will add difficulty to the task of waste classification [1]. Accurate waste classification plays a fundamental role in supporting effective waste management. However, one of the key challenges in this process is the proper categorization of waste based on its type, particularly in distinguishing between organic and inorganic waste [2]. Organic waste can decompose naturally through biological processes, whereas inorganic waste takes longer, even years, to decompose [3]. These data indicate that inorganic waste management requires further treatment, as most of this material is inherently difficult to decompose.

As mentioned, it is important to pay attention to waste management, especially inorganic waste. Waste management that still relies on manual methods in the waste separation process is less efficient and not environmentally friendly [4]. In Indonesia, people still struggle to classify types of waste, often mixing them together [3]. Waste sorting is a fundamental step necessary to enable efficient and cost-effective recycling [5]. Recycling is an essential approach to reducing the amount of waste sent to landfills, oceans, and other natural environments [6]. For example, the implementation of a recycling system in China has proven its success in overcoming the waste problem [7].

In this situation, a crucial step is to enhance recycling efficiency through automated sorting processes [4].

To further improve recycling efficiency, automated sorting systems supported by deep learning techniques have become increasingly important in enabling more accurate and efficient waste classification [8], [9]. An architecture in automatic classification is the Convolutional Neural Network (CNN). With its unique ability to recognize and classify visuals, CNN is recognized as one of the most dominant approaches in image recognition and is the leading choice in this field [8]. CNN is specialized artificial neural networks developed to process complex data structures and perform various tasks, including image classification, audio classification, and Natural Language Processing (NLP) [10].

CNN has been widely used in previous research to classify waste, as seen in research [11], which employs CNN and Support Vector Machines (SVM) techniques to differentiate plastic bottle waste based on brand. The results show that the CNN has 99% accuracy, which is higher than that of the SVM. Other research by [12] used a CNN to differentiate waste types into organic and inorganic, achieving an accuracy of 98.92%. Furthermore, research conducted by [13] CNN method, when integrated with Multilayer Hybrid feature extraction, was reported to achieve 92.6% accuracy in waste classification.

The continuous evolution of CNNs has significantly contributed to the advancement of deep learning research [14]. Improvements in CNN applications have been achieved through refinements in model architecture and the expansion of network depth [15]. Exemplary architecture is essential for performing effective image classification in various categories [16].

The CNN algorithm has various architectures, including Mobilenetv2 and Densenet169, which are designed to meet different needs [17]. MobileNetV2 is an efficient lightweight neural network model that is well suited for deployment on resource-constrained devices with limited memory and computational capabilities [18], [19]. The results of this study show that MobileNetV2 attained an accuracy of 82.92% in the four-class classification of domestic waste, indicating robust model performance and outperforming the other CNN architectures examined. DenseNet169 is a CNN architecture characterized by dense connectivity, where each layer receives input from all preceding layers, thereby facilitating effective feature reuse and improved information transmission throughout the network [20]. Through this structure, DenseNet169 can reduce parameter redundancy, strengthen feature propagation, and improve learning efficiency during model training [21].

In addition, research by [22] also tested Mobilenetv2 to classify plastic waste. In their study, Mobilenetv2 achieved 86% accuracy using the WaDaBa dataset, which covers four types of plastic. These results show that Mobilenetv2 can help manage plastic waste more effectively and efficiently.

Although Mobilenetv2 is quite good at garbage classification, other studies with different CNN architectures also show significant results. For example, research [21] created the PublicGarbageNet framework for classifying public waste. They used a new dataset comprising 10,624 images, divided into four main categories and ten subcategories of trash. The evaluation of multiple CNN architectures demonstrated that DenseNet169 was the most effective model, achieving an accuracy of 96.35%. The results further indicate that DenseNet169 performs better than the other tested architectures in both predictive accuracy and processing speed. Then, another study by [23] utilized DenseNet169 to classify waste images, achieving 82.8% accuracy. This result demonstrates that DenseNet169 outperforms other models, such as AlexNet, GoogLeNet, and VGG.

The evolution of CNN architectures has led to the introduction of EfficientNet, which applies a compound scaling strategy to proportionally optimize network depth, width, and image resolution. Relative to established models such as ResNet, MobileNet, and DenseNet, EfficientNet has been reported to achieve stronger classification accuracy while maintaining higher computational efficiency across diverse image classification benchmarks. Its inclusion in waste classification research could provide further performance gains, especially in resource-constrained environments [24]. Although this study focuses on MobileNetV2 and DenseNet169, consideration of EfficientNet in future work may enhance the overall system design.

Despite substantial advancements in CNN-based waste classification, relatively few studies have comprehensively explored the role of image background characteristics in influencing model classification performance. In practical environments, waste images are commonly captured under diverse and visually cluttered conditions. Background elements can either help by providing contextual clues or hinder by introducing noise. From a theoretical standpoint, the concept of contextual learning suggests that CNNs do not merely focus on the primary object but also learn from surrounding visual cues. Studies in scene understanding and image classification have shown that background information can significantly influence model predictions, particularly in tasks with high intra-class similarity. While some may expect background to assist classification, no prior work has quantitatively compared the performance of CNN models with different architectural complexities on datasets with and without background.

This research fills this research gap by systematically comparing two CNN architectures with different levels of complexity, namely MobileNetV2 and

DenseNet169, using datasets derived from the same source but prepared in two variants: with background context and without background context. The novelty of this research lies in analyzing how background information affects classification performance, depending on the model's architectural depth and its capacity for learning contextual features. By controlling the dataset source and applying consistent preprocessing, this study effectively isolates the role of background in classification accuracy. The findings contribute not only to model selection strategies but also to a deeper understanding of contextual learning behavior in CNN-based waste classification systems.

To further explore whether background elements serve as helpful contextual cues or introduce classification bias, this study is framed around the hypothesis that model complexity influences sensitivity to contextual features. Specifically, we hypothesize that deeper CNN architectures, such as DenseNet169, can exploit background information more effectively, while lightweight models like MobileNetV2 may derive less benefit or even be prone to overfitting.

2. Methods

This chapter outlines the methodological approach employed to assess the influence of image background on CNN-based waste classification, including data collection, data labeling, dataset partitioning, augmentation, model architecture design, and evaluation procedures.

2.1 Data Collection

The dataset used in this study was the publicly accessible "Garbage Classification" dataset from Kaggle, which contains 2,527 recyclable waste images organized into six classes: plastic, metal, glass, cardboard, trash, and paper. This dataset is available via the DOI https://doi.org/10.34740/kaggle/ds/81794. The dataset comprises 482 images of plastic, 410 images of metal, 501 images of glass, 410 images of cardboard, 130 images of trash, and 594 images of paper.

To assess how background context influences classification performance, the original dataset was duplicated and further processed by removing background components using the U²-Net (U2-Net) model, a deep learning-based architecture designed for salient object detection [25]. This study employed the pretrained U2Net.pth model obtained from the official GitHub repository (https://github.com/xuebinqin/U-2-Net). During the segmentation stage, each image was resized to 320 × 320 pixels, converted into a grayscale saliency map, and thresholded at 0.5 to produce a binary mask. This mask was then used to isolate the foreground waste object, while the background region was replaced with a uniform white color using RGB values of 255, 255, and 255 to maintain visual consistency. Finally, all processed images were resized to 224 × 224 pixels in accordance with the input requirements of the CNN architectures used in this study. No manual correction or morphological post-processing was applied, allowing the preprocessing workflow to remain fully automated and reproducible.

This preprocessing produced two dataset variants: one with the original background and one with the background removed. To maintain a fair comparison, both dataset variants were divided using the same training, validation, and testing proportions of 80%, 10%, and 10%, respectively. The data split was generated with a fixed random seed of 42 using the `split-folders` Python package. In total, the final dataset consisted of 5,054 images, with each variant containing 2,527 images.

The decision to use six general waste categories was based on the labeling structure provided in the publicly available dataset. These categories are widely used in prior waste classification research and represent high-level distinctions that are practically relevant for real-world sorting systems. More granular categorization (e.g., subtypes of plastic or paper) was not applied due to limited data availability and potential class imbalance, which could bias model training and compromise the study's primary focus—analyzing the effect of image background on model performance.

2.2 Data Labelling

At this stage, the Kaggle dataset had been pre-annotated by the dataset provider and categorized into six waste classes: cardboard, glass, metal, paper, plastic, and trash. Consequently, the dataset could be used directly for model training without requiring any additional labeling process. Figure 1 provides visual examples of waste images in their original form with background, while Figure 2 shows corresponding examples after background removal using U²-Net [26]. Each figure includes samples from different waste categories to illustrate the preprocessing differences between the two dataset variants.

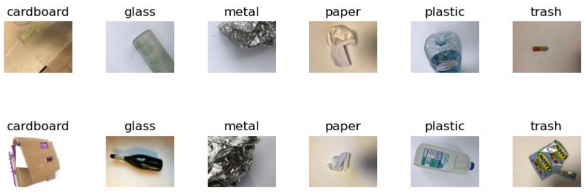

Figure 1. Examples of waste background images

Figure 2. Waste image examples after background removal

2.3. Data Sharing

In this phase, the annotated dataset was partitioned into training, validation, and testing subsets. The training subset was employed to fit the CNN model, the validation subset was used to supervise and validate the learning process, while the testing subset served to evaluate the model's final classification performance. This division is performed using the split-folders module in Python, with a particular seed [27], and follows the general ratio of 8:1:1 [28]. The details of the dataset-sharing settings using this module are provided in Table 1.

Table 1. Splitolders Configuration

Parameter	Mark
Division Type	Ratio
Seeds	42
Training Ratio	8
Validation Ratio	1
Testing Ratio	1

After dividing the data according to the configuration in Table 1, three datasets were produced: training, validation, and testing. The details of this data are explained in Table 2.

Table 2. Number of images in each training, validation, and testing class

Model	Training Data	Data Validation	Testing Data
Cardboard	322	40	41
Glass	400	50	51
Metal	328	41	41
Paper	475	59	60
Plastic	385	48	49
Trash	109	13	15
Total	2019	251	257

2.4 Data Augmentation

To improve the variability and representational richness of the training data, data augmentation was applied through a series of random image transformations. Prior to this process, all images were resized to 224 × 224 pixels to ensure compatibility with the input specifications of the CNN models. The augmentation process included shear transformation, zooming, horizontal flipping, rotation, height shifting, width shifting, and brightness adjustment. Each image in the training set was augmented once, effectively doubling the training set size from 2,019 to 4,038 images. Augmentation was applied uniformly across all six waste classes to maintain class balance and prevent bias.

Alongside data augmentation, dropout regularization and early stopping strategies were incorporated during model training to mitigate overfitting and improve the model's generalization capability. The model's generalization capability was monitored through its performance on the validation and testing sets. These combined strategies ensured that the relatively small dataset could still support robust and reliable model training. Figure 3 illustrates various data augmentation techniques applied to images from the metal category, simulating different real-world visual conditions.

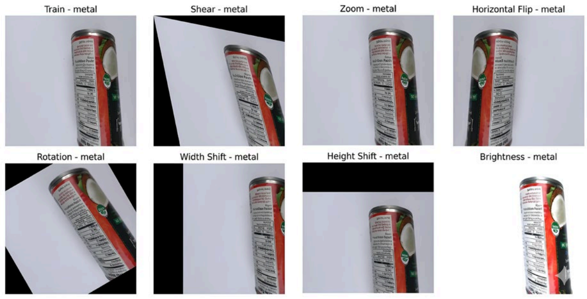

Figure 3. Example of Metal Label Augmentation

These transformations were carefully chosen to simulate common visual variations in real-world waste images. For instance, rotation and flipping reflect different object orientations, brightness adjustment simulates lighting changes, and shifting or zooming mimics differences in object scale and positioning.

The application of data augmentation exposed the CNN architectures to a wider variety of visual patterns, thereby improving model generalization. This was reflected in the training curves, which showed decreased validation loss and reduced discrepancies between training and validation performance, indicating better control of overfitting. Such improvement was especially relevant for models trained using background-free images, as these datasets may contain less contextual information.

2.5 Classification Architecture for Images

Image-based waste recognition plays a significant role in advancing waste management and recycling systems [6]. This process focuses on detecting and classifying different waste types from image data, which can enhance operational efficiency in waste sorting and automated recycling applications [29]. Accordingly, this study applies two CNN architectures, MobileNetV2 and DenseNet169, to assess their performance in classifying waste images according to predetermined categories.

Figure 4 shows the MobileNetV2 architecture, which is a compact and efficient convolutional neural network model specifically designed to support implementation on mobile and resource-constrained devices [6]. The model employs an inverted residual bottleneck design integrated with depthwise separable convolutions, allowing computational complexity to be substantially reduced while maintaining classification accuracy. This architectural design enables MobileNetV2 to achieve high image classification performance while maintaining fast inference speeds [29-31]. MobileNetV2 is suitable for applications on devices with limited resources [18].

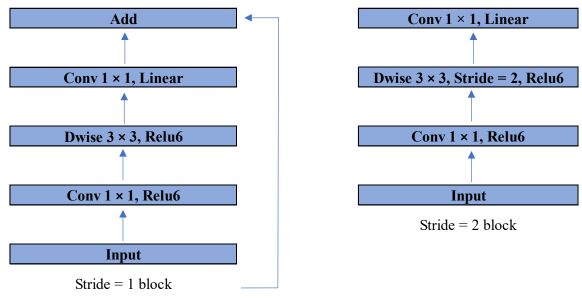

Figure 4. Architectural Structure of MobileNetV2 [25]

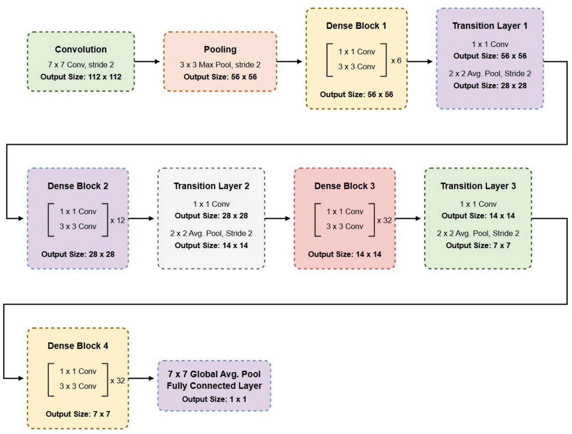

Figure 5. DenseNet169 architecture [32]

Figure 5 illustrates the architecture of DenseNet169, a deep CNN that enhances information flow by establishing direct connections between all layers in a feed-forward manner [33], [34]. This design facilitates efficient gradient propagation during training and encourages feature reuse, which contributes to improved performance in image classification tasks [21]. By enabling deeper feature learning, DenseNet169 achieves high accuracy while maintaining computational efficiency. Due to these advantages, it is particularly suitable for applications requiring robust feature extraction and high classification precision [6], [35].

2.6 Evaluation

In this phase, CNN model performance was evaluated using two complementary metrics: the confusion matrix and the Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). These methods were applied to obtain a more rigorous and comprehensive assessment of the models' classification ability. The confusion matrix served to evaluate classification performance by contrasting the predicted outputs with the actual class labels [23]. It was constructed from the model evaluation results on the dataset that had been divided into 80% training data, 10% validation data, and 10% testing data. This matrix enabled a systematic comparison between predicted classes and ground-truth labels [36]. The confusion matrix structure is presented in Table 3.

Table 3. Confusion Matrix [37]

	True	False
True	TP Correct result	FP Unexpected result
False	FN Missing result	TN Correct absence of result

True Positive (TP) represents the condition in which a model correctly predicts an instance as positive, while True Negative (TN) denotes the correct classification of an instance as negative. By contrast, False Positive (FP) refers to the misclassification of a negative instance as positive, whereas False Negative (FN) describes the misclassification of a positive instance as negative. Based on the confusion matrix, several key evaluation measures can be computed to assess classification performance, namely accuracy, precision, recall, and F1-score, as presented in Equations 1–4.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

$$F1 - score = 2 \times \frac{Recall \times Presisi}{Recall + Presisi} \quad (4)$$

The Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) is an evaluation measure that indicates the ability of a classification model to discriminate between positive and negative classes within a dataset [38]. Following the training process, the model produces probability estimates for the testing data, which are subsequently used to construct the ROC curve. In this study, the ROC-AUC curve was generated by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) using the roc_curve function from the sklearn library [39]. The parameters applied in this process were y_true, representing the actual class labels of the testing data, and y_score, representing the predicted probability scores. Accordingly, the ROC-AUC curve serves as an important evaluation tool for measuring classification performance while analyzing the trade-off between model sensitivity and specificity [38].

3. Results and Discussions

This research aims to determine whether the presence of a background can increase the accuracy of image classification. Therefore, two datasets are used: one with a background and one without, to examine the difference in their influence on classification accuracy. This research will use two different CNN architectures, MobileNetV2 and DenseNet169, to compare their performance in improving image classification accuracy.

These two models will be trained using predetermined parameters, as shown in Table 4.

Table 4. Configuration of CNN Model Parameters.

Parameter	MobileNetV2	DenseNet169
Include_top	False	False
Weights	Imagenet	Imagenet
Number of neurons	64	64
Dense layer activation	Relu	Relu
Optimizer	Adam	Adam
Learning Rate	0.001	0.001
Batch size	32	32
Epochs	20	20

The selected parameters provide a comprehensive description of the model training configuration. To evaluate the proposed model, the confusion matrix and ROC-AUC curve were analyzed as complementary performance assessment tools. This evaluation enables a clearer understanding of the model's capability in classifying waste images and helps identify aspects that require further optimization.

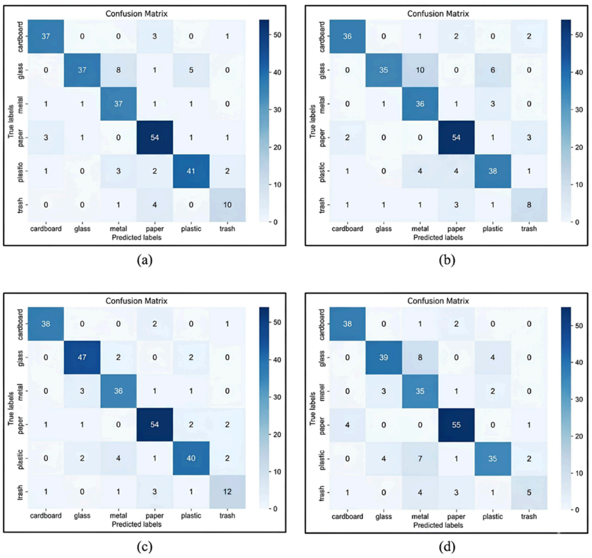

Figure 6. Confusion Matrix Results

Figures 6 (a) MobileNetV2 Dataset Background, (b) MobileNetV2 Dataset Non-Background, (c) DenseNet169 Dataset Background, (d) DenseNet169 Dataset Non-Background display the results of the confusion matrices from two datasets applied with the MobileNetV2 and DenseNet169 algorithms. Table 5 presents a comparative summary of F1-scores across six waste categories, evaluated using two CNN architectures, MobileNetV2 and DenseNet169, on datasets with and without background. This table provides a clear view of how the presence or absence of background information influences model performance at the class level.

Table 5. F1-Score Comparison by Class

Class	MobiNet V2 (BG)	MobiNetV2 (No BG)	DenseNet 169 (BG)	DenseNet1 69 (No BG)
Cardboard	89.16%	90.00%	93.83%	90.48%
Glass	82.22%	76.09%	90.38%	79.59%
Metal	82.22%	77.42%	85.71%	72.92%
Paper	86.40%	88.52%	91.53%	90.16%
Plastic	84.54%	77.55%	84.21%	76.92%
Trash	68.97%	55.17%	75.00%	43.48%

As shown in Table 5, the DenseNet169 model generally achieves higher F1-scores than MobileNetV2, exceptionally when trained on the background-inclusive dataset. These results suggest that deeper models, such as DenseNet169, can better leverage contextual background cues to enhance classification accuracy, especially for visually ambiguous classes like trash and glass. The F1-score values were derived from confusion matrices and classification reports, offering a balanced measure of precision and recall across all experimental conditions.

Table 6. Comparison of Results with Background Dataset.

Methods	Accuracy	Precision	Recall	F1-Score
MobileNetV2	0.8405	0.8307	0.8223	0.8225
DenseNet169	0.8833	0.8634	0.8738	0.8678

Table 7. Comparison of Results with Non-Background Datasets.

Methods	Accuracy	Precision	Recall	F1-Score
MobileNetV2	0.8054	0.7812	0.7752	0.7746
DenseNet169	0.8054	0.7826	0.7516	0.7559

Tables 6 and 7 compare the performance of MobileNetV2 and DenseNet169 in terms of accuracy, precision, recall, and F1-score using both background and non-background image datasets. The highest classification accuracy was 88.33%, achieved by the DenseNet169 model on the background dataset. These findings suggest that the evaluated models demonstrated generally robust performance across multiple evaluation metrics and dataset conditions.

Figure 7 presents the accuracy, precision, recall, and F1-score obtained from a single train-validation-test partition. While these findings suggest promising classification performance, particularly for the DenseNet169 model trained on background images, the results should be interpreted with caution due to their dependence on a single data split. Future experiments should incorporate multiple runs with varying random seeds and report both the average scores and standard deviations. Such an approach would strengthen the robustness of the evaluation and provide a more reliable indication of model consistency and generalizability.

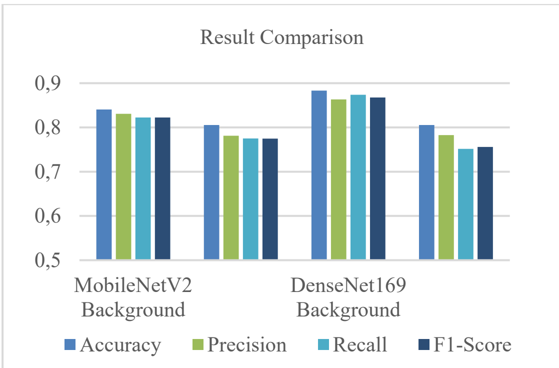

Figure 7. Results Comparison

The subsequent analysis examines the class-level recognition capability of the evaluated models. The

ROC-AUC curve is used to assess how effectively each model discriminates between waste categories. For clearer interpretation, the ROC-AUC results shown in Figures 8a–8d are converted into tabular form, allowing the AUC values for each waste type identified by MobileNetV2 and DenseNet169 to be compared more systematically.

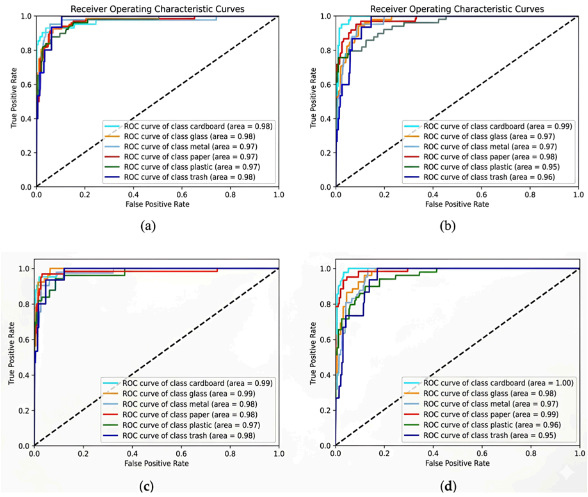

Figure 8. ROC-AUC Curve Results

The discrepancy between AUC and accuracy arises from the different aspects of model performance captured by these metrics. Accuracy reflects the percentage of samples assigned to the correct class, whereas AUC evaluates the model's discriminatory ability by measuring how well it ranks class probabilities. In multiclass classification tasks, particularly when the dataset is imbalanced, a model may obtain only moderate accuracy but still demonstrate strong separability between classes, which is represented by relatively high AUC scores.

In this research, the macro-averaged AUC was calculated using a one-vs-rest (OvR) framework for all six waste classes. Under this procedure, each class is individually treated as the positive class, while all remaining classes are grouped as negative. This approach supports a fairer class-level comparison because it reduces the influence of unequal class distributions. Consequently, higher AUC scores suggest that the models can assign relatively higher probability rankings to the correct class, even when the final predicted labels are not always accurate. As reported in Table 8, the DenseNet169 model trained with background images shows the most consistent and superior performance in discriminating among waste categories compared with the other model configurations. This result aligns with prior research [39], which confirms that the higher the ROCAUC value approaches 1, the better the model is at distinguishing between classes.

Table 8. ROC-AUC Model Curve Results.

Method	Cardboard	Glass	Metal	Paper	Plastic	Trash
Mobilenetv2 Backgrounds	0.98	0.98	0.97	0.97	0.97	0.98
Mobilenetv2 Non-background	0.99	0.97	0.97	0.98	0.95	0.96
Densenet169 Backgrounds	0.99	0.99	0.98	0.98	0.97	0.98
Densenet169 Non-background	1.00	0.98	0.97	0.99	0.96	0.95

The results from both evaluation methods indicate that the DenseNet169 model using the background dataset produced the best overall performance among the evaluated models. The confusion matrix confirms this result, with the model attaining an accuracy of 88.33%. In addition, the ROC-AUC analysis demonstrates strong discriminatory capability across all waste categories, with AUC scores of 0.99 for cardboard, 0.99 for glass, 0.98 for metal, 0.98 for paper, 0.97 for plastic, and 0.98 for trash.

This superior performance is closely linked to the inherent strengths of the DenseNet169 architecture. DenseNet169 uses dense connections, where each layer receives input from all preceding layers. This facilitates better feature reuse and gradient flow, allowing the network to learn richer representations, especially in tasks involving subtle differences between visual classes. Compared with MobileNetV2, which prioritizes efficiency and compactness through depthwise separable convolutions and inverted residuals, DenseNet169 can better capture complex patterns in high-dimensional image data.

Moreover, considering that waste images, especially those with a background, often include fine-grained textures, lighting variations, and overlapping visual elements, DenseNet's ability to preserve and integrate hierarchical features across layers becomes a distinct advantage. While optimized for speed and low computational cost, MobileNetV2 may lose critical spatial detail due to its lightweight design [6]. The trade-off inherent in these architectural designs helps account for the differences in model performance, especially with respect to accuracy and F1-score.

These findings further reinforced by the overall evaluation of background usage. In testing two Convolutional Neural Network architectures, MobileNetV2 and DenseNet169, the use of background datasets was shown to improve model accuracy in classifying waste. The test results confirm that the DenseNet169 model with background data achieved the highest accuracy, with an improvement of approximately 8% compared with its non-background counterpart. The diagram in Figure 9 clearly illustrates that incorporating background consistently enhances the classification performance of both models.

In addition to accuracy, background removal has potential implications for computational efficiency. Images without a background typically contain less visual complexity and noise, thereby reducing the computational burden during feature extraction. This simplification may benefit lightweight models, such as MobileNetV2, in terms of faster training and inference. However, in this study, computational performance metrics such as training time and memory usage were not systematically recorded. Therefore, further background investigation is needed to quantitatively evaluate whether background removal contributes to efficiency gains without sacrificing classification performance.

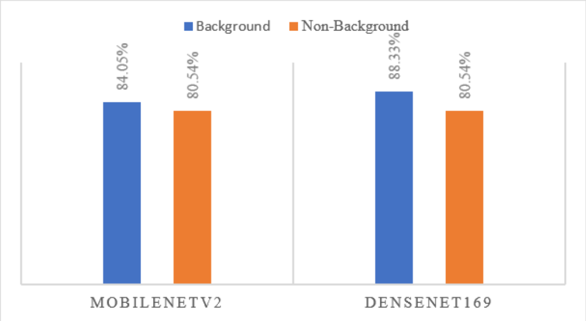

Figure 9. Comparison of the accuracy of MobileNetV2 and DenseNet169 on background and non-background datasets

The success of DenseNet169 with a background dataset in this research is not only visible from its high accuracy rate in Figure 9, but also supported by its unique architectural design. The DenseNet169 architecture uses information from various complex features through dense connections between its constituent layers [34]. Additionally, the presence of background information in the dataset aids the classification process.

The background information in the dataset provides additional context to increase the accuracy of waste type identification [40]. The background can provide clues about the environment or visual texture to support the model in distinguishing waste types [41]. While removing the background can theoretically encourage the model to concentrate on the main visual characteristics of the waste object, this process may also discard valuable contextual information that contributes to more accurate classification [42].

Accordingly, retaining image backgrounds in the dataset can provide supplementary contextual benefits for garbage classification in both DenseNet169 and MobileNetV2 architectures. This finding is strengthened through comparison with prior studies that applied DenseNet169, as presented in Table 9.

Table 9. Comparison of Research Results with Previous Research.

Study	Number of Datasets	Class	Accuracy
This research	2527	6	88.33%
(Zeng et al., 2020)	10224	10	96.35%
(Zhang et al., 2021)	18911	5	82.8%
(Ahmed et al., 2023)	5000	5	94.4%

Based on the results shown in Table 8, the findings suggest that the model used in this study achieved an accuracy level that aligns with those reported in previous studies. Although it does not achieve the highest accuracy as research by [19], which reported an accuracy of 96.35%, the accuracy results obtained are still competitive and in line with previous research.

Prior studies have consistently shown that the size and diversity of datasets, as well as the number of waste categories, significantly influence model performance. Larger datasets with more granular class distinctions facilitate better generalization by exposing the model to a broader range of features. However, factors such as preprocessing quality, parameter tuning, and overfitting control also play critical roles in determining the accuracy of the final results.

Beyond quantitative metrics, the study's findings offer practical implications for real-world deployment. The superior performance of DenseNet169 on background-inclusive images suggests its suitability for integration into automated waste management systems, such as smart recycling stations or conveyor-based sorting lines equipped with camera sensors. Such systems could benefit from the model’s capacity to perform high-precision waste classification in real time, thereby improving overall operational efficiency. Meanwhile, despite its comparatively lower accuracy, MobileNetV2 offers notable computational advantages, making it appropriate for deployment in mobile or embedded environments, including portable waste scanners and IoT-enabled smart bins for distributed waste monitoring. This observation aligns with recent studies [43-45] that emphasize the trade-off between helpful contextual cues and harmful background bias in CNN-based visual classification. These works highlight that while context can assist recognition, it may also introduce shortcuts that reduce model generalization, particularly in transfer learning scenarios. Future research should further investigate model robustness under heterogeneous environmental settings, including changes in illumination, image resolution, background clutter, and the presence of motion blur. Increasing dataset diversity by incorporating a wider range of waste categories and assessing newer CNN architectures, including EfficientNet and hybrid models, could further improve classification accuracy, robustness, and scalability.

It is important to acknowledge that the evaluation outcomes presented in this study were based on only one train-validation-test split, with proportions of 80%, 10%, and 10%, respectively, generated using a fixed random seed. While this setup allows for a controlled comparison between model variants under consistent conditions, it does not account for variability introduced by different data partitions. Therefore, the observed differences in performance should be interpreted with caution. As highlighted in recent [46], even minor variations in dataset composition can lead to performance fluctuations, especially when the model is sensitive to background features. To strengthen the reliability of the findings, subsequent experiments should include multiple training iterations using different random seeds and report the averaged results accompanied by standard deviation values. Moreover, future research should explore more diverse and larger datasets better to reflect real-world variability in waste types and imaging conditions. This includes testing model robustness under cross-domain scenarios, such as using images from different geographic regions, lighting environments, or camera types, which would assess generalization beyond the training domain. Evaluating the impact of synthetic data augmentation and semi-supervised learning approaches may also provide scalable strategies for improving performance in data-limited situations. Such directions strengthen the applicability of deep learning models in practical, heterogeneous deployment environments.

4. Conclusions

The findings of this study indicate that the inclusion of background information in image datasets contributes meaningfully to improving deep learning-based waste classification performance. Among the evaluated models, DenseNet169 trained on the background dataset produced the best result, achieving an accuracy of 88.33% and outperforming both MobileNetV2 and the models trained on background-free datasets. This result emphasizes the role of contextual visual cues in supporting more accurate classification. The superior performance of DenseNet169 can be attributed to its dense connection mechanism, which facilitates richer feature propagation and more effective representation learning. Moreover, its high precision, recall, and F1-score demonstrate its potential applicability in real-world intelligent waste management systems.

Further research could explore the effect of additional background diversity on classification across broader waste categories and more complex datasets. Expanding the dataset with more varied waste types and refining models with advanced data augmentation techniques could yield more accurate and robust systems. Moreover, testing alternative architectures and hybrid models may enhance classification efficiency and scalability in resource-constrained environments. While this study demonstrates that background-inclusive images enhance classification performance, the improvement may be partially due to the model's reliance on contextual cues rather than genu ine object recognition. Since no specific experiments, such as background randomization, saliency mapping, or counterfactual testing, were conducted to isolate or visualize this potential bias, the findings should be interpreted with caution. Acknowledging this limitation is essential to avoid overgeneralization and guide future work toward more interpretable and generalizable waste classification models.

Acknowledgements

The authors confirm that they have no competing interests to declare.

References

[1] Z. Kang, J. Yang, G. Li, and Z. Zhang, "An Automatic Garbage Classification System Based on Deep Learning," *IEEE Access*, vol. 8, pp. 140019–140029, 2020, doi: 10.1109/ACCESS.2020.3010496.

[2] Sunardi, A. Yudhana, and M. Fahmi, "SVM-CNN Hybrid Classification for Waste Image Using Morphology and HSV Color Model Image Processing," *Trait. du Signal*, vol. 40, no. 4, pp. 1763–1769, 2023, doi: 10.18280/ts.400446.

[3] F. Hamami, "Garbage classification using convolutional neural network based transfer learning," 2023, p. 20012. doi: 10.1063/5.0128946.

[4] F. Wu and H. Lin, "Effect of transfer learning on the performance of VGGNet-16 and ResNet-50 for the classification of organic and residual waste," *Front. Environ. Sci.*, vol. 10, 2022, doi: 10.3389/fenvs.2022.1043843.

[5] K. Lin *et al.*, "Applying a deep residual network coupling with transfer learning for recyclable waste sorting," *Environ. Sci. Pollut. Res.*, vol. 29, no. 60, pp. 91081–91095, 2022, doi: 10.1007/s11356-022-22167-w.

[6] M. I. B. Ahmed *et al.*, "Deep Learning Approach to Recyclable Products Classification: Towards Sustainable Waste Management," *Sustain.*, vol. 15, no. 14, 2023, doi: 10.3390/su151411138.

[7] A. Sevinç and F. Ozyurt, "Classification of recyclable waste using deep learning architectures," *FIRAT Univ. J. Exp. Comput. Eng.*, vol. 1, no. 3, pp. 122–128, 2022, doi: 10.5505/fujece.2022.83997.

[8] W. L. Mao, W. C. Chen, C. T. Wang, and Y. H. Lin, "Recycling waste classification using optimized convolutional neural network," *Resour. Conserv. Recycl.*, vol. 164, 2021, doi: 10.1016/j.resconrec.2020.105132.

[9] A. H. Vo, L. H. Son, M. T. Vo, and T. Le, "A Novel Framework for Trash Classification Using Deep Transfer Learning," *IEEE Access*, vol. 7, pp. 178631–178639, 2019, doi: 10.1109/ACCESS.2019.2959033.

[10] O. Campesato, *Artificial Intelligence, Machine Learning, and Deep Learning*. Mercury Learning and Information LLC, 2020. [Online]. Available: https://books.google.co.id/books?hl=id&lr=&id=pqnNDwAAQBAJ&oi=fnd&pg=PT16&dq=Artificial+Intelligence,+Machine+Learning+and+Deep+Learning.+Dulles:+Mercury+Learning+and+Information+LLC.&ots=I2SxMON0UU&sig=Js5avan1NBgUuzkgWiBDlMrIHl8&redir_esc=y#v=onepage&q=

[11] A. Fadlil, R. Umar, Sunardi, and A. S. Nugroho, "Comparison of Machine Learning Approach for Waste Bottle Classification," *Emerg. Sci. J.*, vol. 6, no. 5, pp. 1075–1085, 2022, doi: 10.28991/ESJ-2022-06-05-011.

[12] Sunardi, A. Yudhana, and M. Fahmi, "Improving Waste Classification Using Convolutional Neural Networks: An Application of Machine Learning for Effective Environmental Management," *Rev. d'Intelligence Artif.*, vol. 37, no. 4, pp. 845–855, 2023, doi: 10.18280/ria.370404.

[13] C. Shi, C. Tan, T. Wang, and L. Wang, "A waste classification method based on a multilayer hybrid convolution neural network," *Appl. Sci.*, vol. 11, no. 18, 2021, doi: 10.3390/app11188572.

[14] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of image classification algorithms based on convolutional neural networks," 2021, *MDPI*. doi: 10.3390/rs13224712.

[15] F. Nashrullah, S. A. Wibowo, and D. G. Budiman, "Investigasi Parameter Epoch Pada Arsitektur ResNet-50 Untuk Klasifikasi Pornografi," *J. Comput. Electron. Telecommun.*, 2020, doi: 10.52435/complete.v1i1.51.

[16] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, "Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3840–3854, 2020, doi:

10.1109/TCYB.2020.2983860.
[17] A. Mabrouk, R. P. D. Redondo, A. Dahou, M. A. Elaziz, and M. Kayed, "Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks," *Appl. Sci.*, vol. 12, no. 13, 2022, doi: 10.3390/app12136448.
[18] L. Yong, L. Ma, D. Sun, and L. Du, "Application of MobileNetV2 to waste classification," *PLoS One*, vol. 18, no. 3 March, 2023, doi: 10.1371/journal.pone.0282336.
[19] D. Yu, Q. Xu, H. Guo, C. Zhao, Y. Lin, and D. Li, "An Efficient and Lightweight Convolutional Neural Network for Remote Sensing Image Scene Classification," *Sensors*, 2020, doi: 10.3390/s20071999.
[20] G. S. Susanth, L. M. J. Livingston, and L. G. X. A. Livingston, "Garbage Waste Segregation Using Deep Learning Techniques," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1012, no. 1, p. 12040, 2021, doi: 10.1088/1757-899x/1012/1/012040.
[21] M. Zeng, X. Lu, W. Xu, T. Zhou, and Y. Liu, "PublicGarbageNet : A Deep Learning Framework for Public Garbage Classification," in *Proceedings of the 39th Chinese Control Conference*, 2020, doi: 10.23919/CCC50068.2020.9189561.
[22] J. Bobulski and M. Kubanek, "Deep Learning for Plastic Waste Classification System," *Appl. Comput. Intell. Soft Comput.*, vol. 2021, 2021, doi: 10.1155/2021/6626948.
[23] Q. Zhang, Q. Yang, X. Zhang, Q. Bao, J. Su, and X. Liu, "Waste image classification based on transfer learning and convolutional neural network," *Waste Manag.*, vol. 135, pp. 150–157, 2021, doi: 10.1016/j.wasman.2021.08.038.
[24] Risfendra, G. F. Ananda, and H. Setyawan, "Deep Learning-Based Waste Classification with Transfer Learning Using EfficientNet-B0 Model," *J. RESTI*, vol. 8, no. 4, pp. 535–541, 2024, doi: 10.29207/resti.v8i4.5875.
[25] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, p. 107404, 2020, doi: 10.1016/j.patcog.2020.107404.
[26] T. Jahan and A. Rajee, "Image Background Excluder: A Deep Learning Approach to Object Isolation," no. February, 2025, doi: 10.13140/RG.2.2.18436.64644.
[27] J. Filter, "Split-folders 0.5.1," 3 February 2022. [Interactiv]. Available: https://pypi.org/project/split-folders/.
[28] J. Brown, Z. Gharineiat, and N. Raj, "CNN Based Image Classification of Malicious UAVs," *Appl. Sci.*, vol. 13, no. 1, 2023, doi: 10.3390/app13010240.
[29] H. Zheng and Y. Gu, "Encnn-upmws: Waste classification by a CNN ensemble using the UPM weighting strategy," *Electron.*, vol. 10, no. 4, pp. 1–21, 2021, doi: 10.3390/electronics10040427.
[30] Y. Pang, W. Yu, Y. Zhang, C. Xuan, and P. Wu, "Sheep face recognition and classification based on an improved MobilenetV2 neural network," *Int. J. Adv. Robot. Syst.*, vol. 20, no. 1, 2023, doi: 10.1177/17298806231152969.
[31] E. Amoako, Y. Sheng, M. Khalid, and M. Bock, "Application of SSD-MobileNetV2 for automated defect detection in masonry bridges using AI and IoT," *Adv. Bridg. Eng.*, vol. 6, no. 1, 2025, doi: 10.1186/s43251-025-00179-z.
[32] R. P. Narmadha, N. Sengottaiyan, and R. J. Kavitha, "Deep transfer learning based rice plant disease detection model," *Intell. Autom. Soft Comput.*, vol. 31, no. 2, pp. 1257–1271, 2022, doi: 10.32604/iasc.2022.020679.
[33] T. Chauhan, H. Palivela, and S. Tiwari, "Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, 2021, doi: 10.1016/j.jjimei.2021.100020.
[34] B. A. S. Al-rimy, F. Saeed, M. Al-Sarem, A. M. Albarrak, and S. N. Qasem, "An Adaptive Early Stopping Technique for DenseNet169-Based Knee Osteoarthritis Detection Model," *Diagnostics*, vol. 13, no. 11, 2023, doi: 10.3390/diagnostics13111903.
[35] A. Yadav and D. S. Yadav, "Enhancing Oral Squamous Cell Carcinoma Detection: A Transfer Learning Perspective on Histopathological Analysis Using ResNet-18, AlexNet, DenseNet-169, and DenseNet-201 with Cyclic Learning Rate," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 17s, pp. 689–699, 2024, [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/4937
[36] M. S. Amrullah and S. F. Pane, *Analisis Sentimen Masyarakat Terhadap Kebijakan Polisi Tilang Manual Di Indonesia*. Penerbit Buku Pedia, 2023. Available: https://books.google.co.id/books?id=_Uq5EAAAQBAJ&printsec=copyright&hl=id#v=onepage&q&f=false.
[37] F. Rahmad, Y. Suryanto, and K. Ramli, "Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, 2020. doi: 10.1088/1757-899X/879/1/012076.
[38] A. U. Zailani and N. L. Hanun, "Penerapan Algoritma Klasifikasi Random Forest Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera," *Infotech J. Technol. Inf.*, vol. 6, no. 1, pp. 7–14, 2020, doi: 10.37365/jti.v6i1.61.
[39] F. S. Nahm, "Receiver operating characteristic curve: overview and practical use for clinicians," *Korean J. Anesthesiol.*, vol. 75, no. 1, pp. 25–36, 2022, doi: 10.4097/kja.21209.
[40] K. Xiao, L. Engstrom, A. Ilyas, and A. Madry, "Noise or Signal: The Role of Image Backgrounds in Object Recognition," 2020, [Online]. Available: http://arxiv.org/abs/2006.09994
[41] R. Aljundi, K. Kelchtermans, and T. Tuytelaars, "Task-free continual learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2019, pp. 11246–11255. doi: 10.1109/CVPR.2019.01151.
[42] E. H. E. Yasin and C. Kornel, "Evaluating Satellite Image Classification: Exploring Methods and Techniques," 2024. doi: 10.5772/intechopen.1003196
[43] S. T. Erukude, A. Joshi, and L. Shamir, "Identifying Bias in Deep Neural Networks Using Image Transforms," *Computers*, vol. 13, no. 12, 2024, doi: 10.3390/computers13120341.
[44] E. M. Aminoff, S. Baror, E. W. Roginek, and D. D. Leeds, "Contextual associations represented both in neural networks and human behavior," *Sci. Rep.*, vol. 12, no. 1, pp. 1–12, 2022, doi: 10.1038/s41598-022-09451-y.
[45] P. R. A. S. Bassi, S. S. J. Dertkigil, and A. Cavalli, "Improving deep neural network generalization and robustness to background bias via layer-wise relevance propagation optimization," *Nat. Commun.*, vol. 15, no. 1, pp. 1–16, 2024, doi: 10.1038/s41467-023-44371-z.
[46] S. Uhlig, I. Alkhasli, F. Schubert, C. Tschöpe, and M. Wolff, "A review of synthetic and augmented training data for machine learning in ultrasonic non-destructive evaluation," *Ultrasonics*, vol. 134, no. June 2022, 2023, doi: 10.1016/j.ultras.2023.107041.