

Performance Evaluation of a Machine Learning for Image-Based Waste Management Using Convolutional Neural Networks

Dandy Cahyo Purnomo¹, Eka Aprilia Saputri², Ahmad Hasib Satiri³, Luluk Khatimah⁴, Hartono⁵

^{1,5} Mechatronics Engineering Department, Swiss German University, Jalur Sutera Barat No. 15, Tangerang 15143, Indonesia

² Computerized Accounting, Department of Accounting, Semarang State Polytechnic, Semarang 50275, Indonesia

³ Automation Engineering, Faculty of Technology, University State of Jakarta, Jakarta 13220, Indonesia

⁴ Directorate of Planning, Information Systems, and Digital Transformation, Padjadjaran University, West Java 45363, Indonesia

ARTICLE INFO

Article historys:

Received : 10/03/2026

Revised : 06/04/2026

Accepted : 12/05/2026

Keywords:

Convolutional Neural Network;
Machine Learning; Smart Waste
Management; Transfer Learning;
Waste Classification

ABSTRACT

Waste management is still a significant environmental problem, especially in developing countries with high population density. The waste sorting process, which is still mostly done manually, causes low efficiency and consistency of sorting. Previous studies generally focused on binary classification, while research on lightweight CNN architectures for multi-class waste classification, particularly in simulation-based scenarios with per-class performance analysis, remained relatively limited. This study aims to evaluate the performance of the waste image classification model using a lightweight Convolutional Neural Network architecture, namely MobilNetV2 based on the transfer learning approach. The dataset used consists of four categories of waste (metal, plastic, paper, organic) obtained from public datasets. The flow of this research includes the dataset collection stage, preprocessing in the form of resize and rescale, training model and evaluation model using accuracy, confusion matrix, precision, recall, and F1-Score metrics. The test results showed that the model achieved an accuracy of 92.02% with relatively consistent performance in each class. The misclassification that appears is limited and shows a similar pattern between classes. These findings show that MobileNetV2 is able to provide stable classification performance and has the potential to be used as a supporting model in the development of automated waste sorting systems in the future.



This work is licensed under a Creative Commons Attribution 4.0 International License

Corresponding Author:

Dandy Cahyo Purnomo

Mechatronics Engineering Department, Swiss German University, Jalur Sutera Barat No. 15, Tangerang 15143, Indonesia

Email: dandy.purnomo@sgu.ac.id

1. INTRODUCTION

Waste management is one of the environmental problems that is still a challenge in various countries. The world's waste is expected to increase to 3.4 billion tonnes per year by 2050, with most of it generated in developing countries that lack the infrastructure to recycle waste and divert resources to collect it [1]. Indonesia is one of the most densely populated developing countries. Population growth and increased public consumption activities contribute to the increase in the volume of household waste every year [2].

One of the important stages in waste management is the process of sorting by material type, such as organic waste, plastic, paper, and metal, which aims to support recycling and advanced processing

processes more effectively. In practice, the waste sorting process is still mostly done manually. This method requires a lot of time and effort and relies heavily on human precision so that many are still unable to sort waste properly. Technologies such as bins that can automatically separate waste were developed to address this problem [3]. One of the findings of waste image classification provides a new direction for waste classification problems is the use of Convolutional Neural Network (CNN) to train waste image classification models [4]. The model is trained to recognize different types of waste through visual analysis such as color, shape, and texture.

In several studies, deep learning, especially Convolutional Neural Networks, has been used for digital image-based waste classification. Compared to conventional methods that rely on manual sorting, this method has proven to be effective in extracting visual features automatically and improving classification accuracy [5]. As demonstrated by Zheng et al., the integration on lightweight CNN architecture with an attention mechanism is highly effective for achieving accurate object localization, particularly in identify the spatial position of insect pests[1].

In addition, a deep learning approach has also been used in smart and sustainable urban waste management; This shows that the classification of waste based on images is still relevant and constantly evolving [7]. However, some previous research has still concentrated on the use of binary systems or classifications. Meanwhile, studies evaluating the performance of light CNN architecture in simulation-based multi-class waste classification scenarios with performance analysis per class are still relatively limited.

Based on the research gap, this study aims to evaluate the performance of the MobileNetV2 architecture in conducting multi-class classification of image-based waste in a simulation scenario. The evaluation was carried out using accuracy, precision, recall, and F1-score metrics to analyze the consistency and balance of model performance in each waste class. This research is expected to contribute in the form of a deeper understanding of the capabilities and limitations of MobileNetV2 as a lightweight classification model, as well as its potential to be used as a supporting module in the development of intelligent waste management systems in future research.

2. RESEARCH METHOD

This study uses the *applied experimental research* method with the Convolutional Neural Network *model transfer learning approach* for the classification of waste images. This method aims to evaluate the performance of the machine learning model in classifying images into four categories of waste based on test data. The experimental approach was chosen because it allows for controlled measurement of model performance through quantitative metrics as well as prediction error analysis. The results of the model's performance measurement are expected to be able to be implemented as an intelligence component in the automatic waste classification system. The research workflow is illustrated in **Figure 1**. First, the dataset is collected and categorized into four classes: metal, plastic, paper, and organic. The dataset is subsequently partitioned into three subsets: 80% for training, 10% for validation to prevent overfitting, and 10% for testing. Prior to training, preprocessing is applied through resizing and rescaling to standardize image dimensions and normalize pixel values. The MobileNetV2 architecture is then trained on the prepared subsets, after which model evaluation is conducted to assess classification performance.

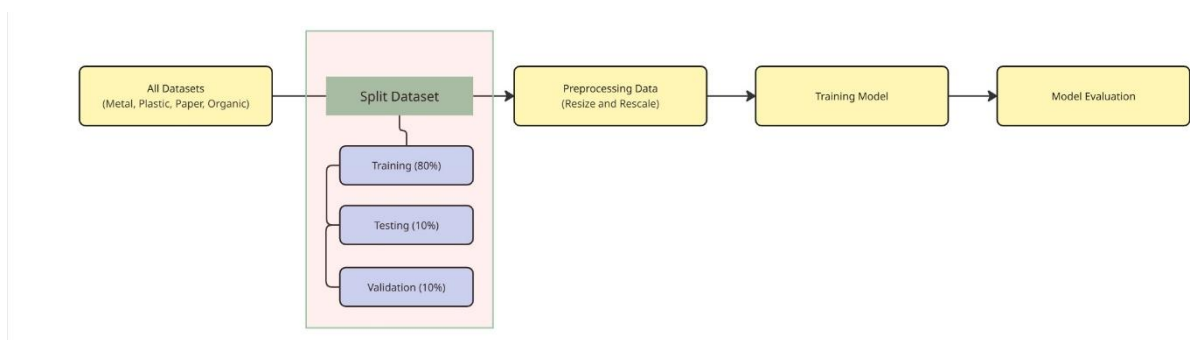


Figure 1. Research workflow

2.1. Dataset Collection

The dataset was taken from the Kaggle public dataset [2][3], which was divided into four categories, namely metal, organic, paper and plastic with a total of 410 in each class. The images are taken from various angles with diverse backgrounds. The dataset is then stratified into 80% for data training, 10% for testing, and 10% validation. Table 1 shows the distribution of the dataset ratios used. An example of the dataset used is in “Figure 2”. Due to the limitations of the dataset, an augmentation technique was carried out to increase the variety of the dataset, namely in the form of random horizontal or vertical reversals with a rotation of 15° [4].

Table 1. Dataset distribution

	Total	Train (80%)	Test (10%)	Val (10%)
Metal	410	328	41	41
Organic	410	328	41	41
Paper	410	328	41	41
Plastic	410	328	41	41
	1.640	1.312	164	164

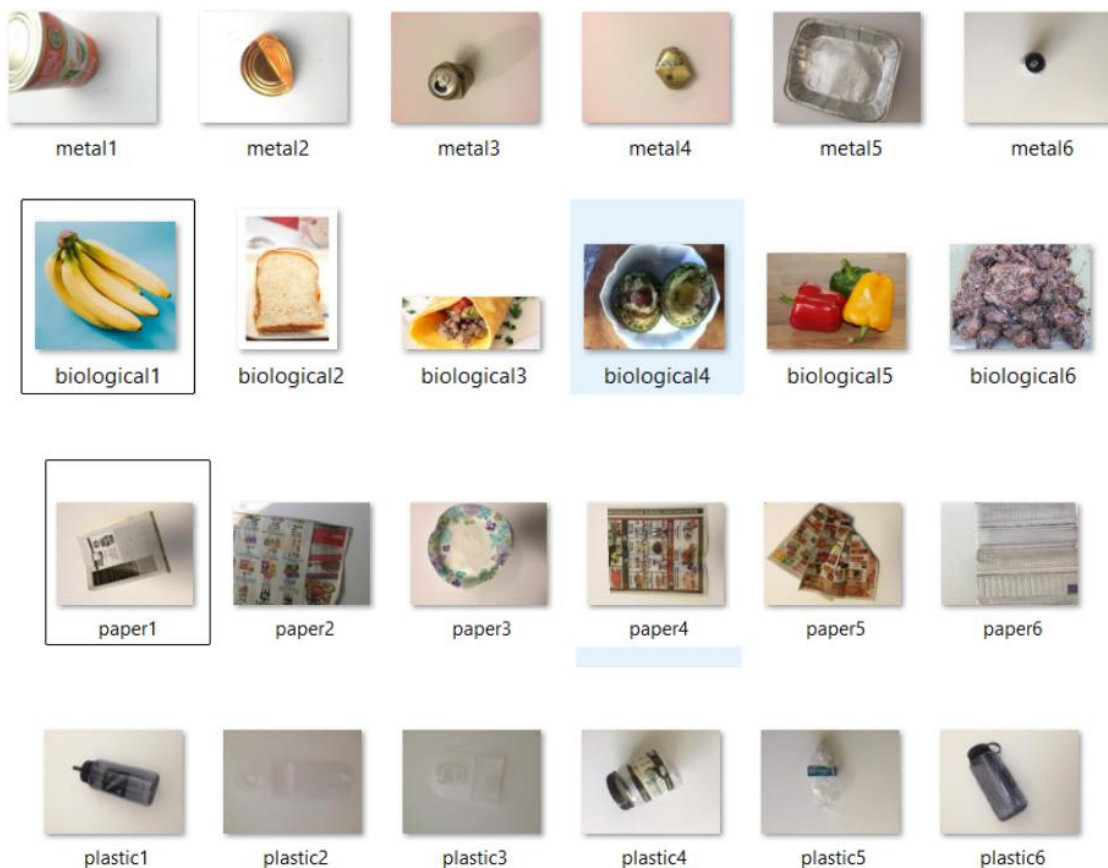


Figure 2. Example of the dataset used in the study is divided into four classes, namely metal, organic, paper and plastic

2.2. Preprocessing Data

Image preprocessing is done by resize the entire image size to 224 x 224 pixels to meet the input layer specifications of the MobilNetV2 architecture. This process is important in the data preprocessing stage because it ensures that the neural network can process each image consistently without changing

its key visual features [5]. Furthermore, the pixel value is normalized to a floating value of 0 to 1 to accelerate the convergence of the training process [6].

2.3. Training Model

The model runs on the MobilNetV2 architecture. The MobilNetV2 was chosen because it is lightweight and can balance performance and computing efficiency [7]. The MobileNetV2 model is initialized from ImageNet, which is transfer learning so that the model can leverage knowledge from big data such as ImageNet to improve image classification tasks [13].

The model was trained using the Adam optimizer with a learning rate of 0.0001, a batch size of 32, and a number of epochs of 20. The relatively small learning rate selection aims to ensure that the weight update process takes place stably at the classification layer. During the training process, data augmentation is applied to the training data to improve visual variation of the image. This augmentation is based on the ImageDataGenerator class provided by the Keras deep learning neural network library [9][10]. Meanwhile, validation data is used to monitor model performance and detect potential overfitting through the observation of accuracy and loss values.

2.4. Model Evaluation

The evaluation of model performance was carried out using a confusion matrix. The confusion matrix represents the actual value and predicted value of the model [11]. The model is then evaluated by calculating precision, recall, and F1-score metrics for each of its classes defined in Equations (1) – (4). The main components of the confusion matrix consist of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). The four components are calculated for each class where one class is treated as a positive class and the other class as a negative class. As an illustration, in the paper class, the TP value represents the number of paper images that are correctly predicted to be papers. The FP value indicates the number of images from other classes (e.g. plastic or metal) that are incorrectly predicted to be paper. The FN value represents the number of paper images that are incorrectly classified as other classes, while TN represents the number of images from other classes that are correctly predicted as non-paper. The same approach is applied to the metal, organic, and plastic classes.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

It is used to measure the level of reliability of a model's prediction of a class, i.e. the extent to which the resulting positive prediction is correct.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

To measure the model's ability to detect all actual samples of a class.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (3)$$

To provide a measure of balance between precision and recall in order to be able to represent model performance in more detail.

3. RESULTS AND DISCUSSION

The test results on the test data showed that the model achieved an accuracy of 92.02% with a low loss value. These results show that the transfer learning approach with MobileNetV2 is able to classify waste images well even though the model is not fine-tuned. This high performance indicates that the visual features learned from ImageNet are representative enough to distinguish categories of garbage that have clear visual differences.

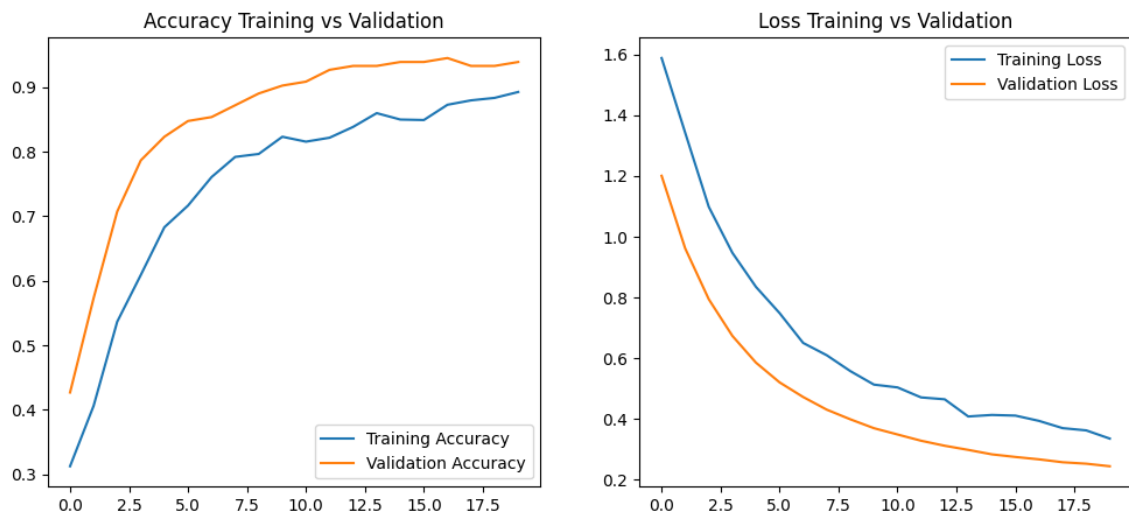


Figure 3. Training or validation loss and accuracy

In “Figure 3,” the model's performance shows an overall improvement trend during the training process, although it is accompanied by fluctuations between epochs. The accuracy of the training data which at the beginning of the training was at 0.2982 increased to a range of 0.85, with some temporary decreases at certain epochs. The accuracy of the validation data also increased gradually and reached a high value of 0.94 in the 17th epoch, but did not increase monotonically throughout the training. The lowest validation loss was recorded at 0.26 in the 17th epoch. The achievement of the highest validation accuracy value in conjunction with the lowest validation loss indicates that the performance of the model at the epoch represents the most optimal learning conditions based on the validation data, according to the pattern shown by the graph and table of training results.

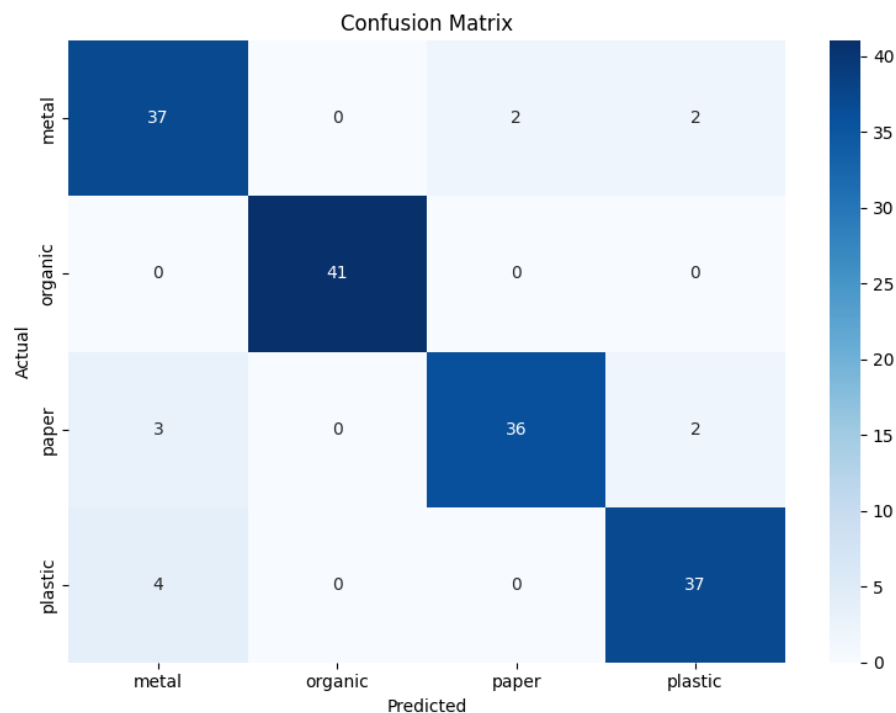


Figure 4. Confusion Matrix

The confusion matrix shows that most of the test data is correctly classified, with a relatively consistent pattern of error between classes. The organic class exhibits the highest consistency, where all 41 data are predicted precisely without errors. In the metal class, out of 41 data, there are 4 prediction errors that are consistently divided into paper (2 data) and plastic (2 data) classes. The paper class

showed 5 prediction errors, with a similar pattern of errors, namely 3 data misclassified as metal and 2 data as plastic. Meanwhile, in the plastic class, out of 41 data, there are 4 data that are consistently incorrectly predicted as metals. A pattern of repeating errors that are concentrated in specific class pairs indicates that the model has stable predictive behavior, where errors are not random, but rather appear in classes with potentially intersecting visual characteristics. Thus, this confusion matrix not only indicates the level of accuracy of the model, but also reflects the consistency of the prediction pattern and the resulting error.

Table 2. Evaluation Matrix in the form of Precision, Recall and F1-Score

	Precision	Recall	F1-Score
Metal	0,8409	0,9024	0,8705
Organic	1,0000	1,0000	1,0000
Paper	0,9474	0,8780	0,9114
Plastic	0,9024	0,9024	0,9024
Macro Avarage	0,9226	0,9207	0,9211
Weighted Average	0,9226	0,9207	0,9211

In the evaluation metrics, the model showed consistent performance across the class. The organic class obtained a precision, recall, and F1-score of 100%, which indicates a perfect level of prediction consistency in the test data so that it is easily recognized by the model. Meanwhile, the metal, paper, and plastic classes also showed varied performance, with F1-score values of 87.05%, 91.13% respectively; and 90.24%. This difference in performance between classes indicates a certain level of visual similarity, especially in the paper and plastic classes. Nonetheless, the F1-Score score obtained still shows a good balance between precision and recall in each class. Overall, the average value of the evaluation metrics reached 92%, which shows that the model is able to consistently classify waste on the dataset used. However, it should be noted that the dataset used was under controlled conditions, so the model's performance under real conditions with variations in lighting, object deformation, and more diverse levels of waste cleanliness could not be fully represented in this study.

In addition to quantitative evaluation, qualitative inference testing was conducted using a Gradio-based interface, as demonstrated in Figures 5 through 8. Figure 5. shows a plastic wrapper sample correctly classified as plastic, Figure 6. depicts a beverage can correctly predicted as metal, Figure 7. presents bread waste classified as organic, and Figure 8. shows a newspaper correctly identified as paper. This test aims to verify the predictive behavior of the model on the input image individually outside of the batch testing scheme. The test results showed that the model was able to provide class predictions that were consistent with the performance patterns obtained in quantitative evaluations, especially in images with clear visual characteristics. These findings indicate that the model not only performs well numerically, but also performs stably in simulated use scenarios, which is relevant for the development of automated waste classification systems in the future.

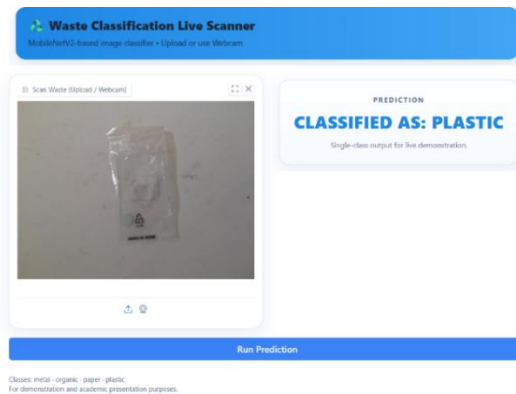


Figure 5. Plastic wrapper waste predicted as plastic

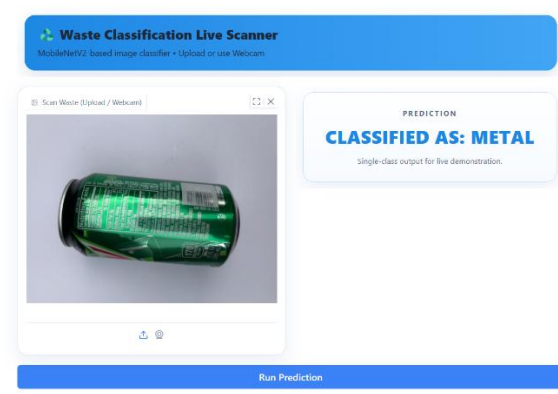


Figure 6. Beverage can waste predicted as metal

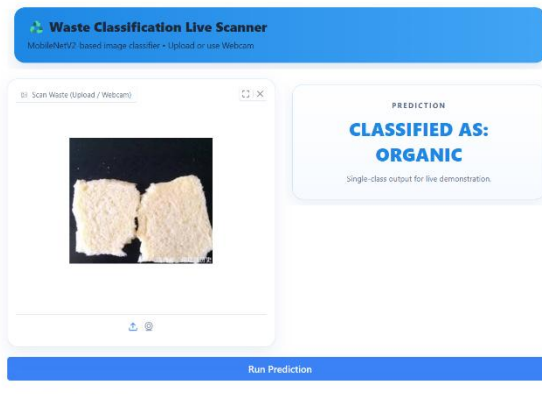


Figure 7. Bread waste predicted as organic

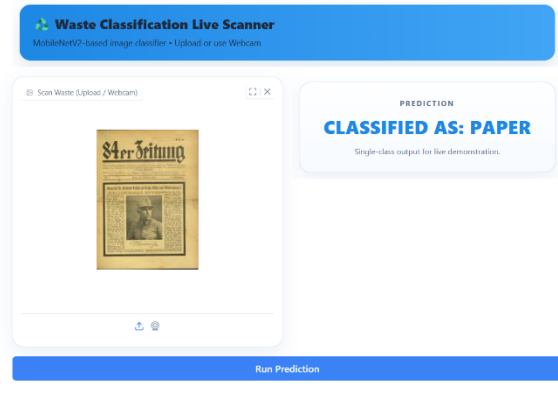


Figure 8. Newspaper waste predicted as paper

4. CONCLUSION

The MobileNetV2 model achieved an accuracy of 92.02% in multi-class waste image classification, demonstrating balanced performance across all categories. Per-class analysis revealed consistent predictive behavior, with misclassification patterns concentrated among visually similar classes. These results confirm that lightweight CNN architectures are capable of delivering stable classification performance, both in terms of overall accuracy and inter-class prediction distribution. The primary contribution of this study lies in the systematic evaluation of per-class performance consistency and misclassification pattern analysis in lightweight models, aspects that have received limited attention in prior literature. From a practical standpoint, the findings suggest that MobileNetV2 holds considerable potential as a classification module in automated waste sorting systems and image-based educational applications. However, validation under real-world environmental conditions and deployment on resource-constrained devices remain important directions for future research.

Acknowledgments

The authors acknowledge the financial support of Swiss German University through internal research funds program.

REFERENCES

- [1] G. Ahmad *et al.*, "Intelligent Waste Sorting for Urban Sustainability Using Deep Learning," *Scientific Reports* 2025 15:1, vol. 15, no. 1, pp. 27078-, Jul. 2025, doi: 10.1038/s41598-025-08461-w.
- [2] "SIPSN - National Waste Management Information System." Accessed: Feb. 21, 2026. [Online]. Available: <https://portal-sipsn.kemnlh.go.id/>

- [3] X. R. Huang, S. S. Yang, W. S. Chen, Y. Q. Zhang, C. Te Lee, and L. B. Chen, "An IoT-Based Smart Trash Cans Monitoring System," in *2021 IEEE 10th Global Conference on Consumer Electronics, GCCE 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 623–624. doi: 10.1109/GCCE53005.2021.9621358.
- [4] Md. W. Rahman, R. Islam, A. Hasan, N. I. Bithi, Md. M. Hasan, and M. M. Rahman, "Intelligent waste management system using deep learning with IoT," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 2072–2087, 2022, doi: <https://doi.org/10.1016/j.jksuci.2020.08.016>.
- [5] N. Nnamoko, J. Barrowclough, and J. Procter, "Solid Waste Image Classification Using Deep Convolutional Neural Network," *Infrastructures (Basel)*, vol. 7, no. 4, 2022, doi: 10.3390/infrastructures7040047.
- [6] T. Zheng, X. Yang, J. Lv, M. Li, S. Wang, and W. Li, "An efficient mobile model for insect image classification in the field pest management," *Engineering Science and Technology, an International Journal*, vol. 39, Mar. 2023, doi: 10.1016/J.JESTCH.2023.101335.
- [7] G. Ahmad *et al.*, "Intelligent waste sorting for urban sustainability using deep learning," *Sci. Rep.*, vol. 15, Jul. 2025, doi: 10.1038/s41598-025-08461-w.
- [8] "Garbage Classification." Accessed: Feb. 10, 2026. [Online]. Available: <https://www.kaggle.com/datasets/hassnainzaidi/garbage-classification>.
- [9] "Garbage Classification EcoSort." Accessed: Feb. 28, 2026. [Online]. Available: https://www.kaggle.com/datasets/itskindastrange/garbage-classification-ecosort?select=garbage_classification.
- [10] H. Yuliansyah, R. A. Saputro, I. I. Khoirunnisa, W. N. S. W. Ali, Y. S. R. Nur, and N. F. M. Radzuan, "Plastic bottle defect detection based on convolutional neural network with MobileNetV2 architecture," *Franklin Open*, vol. 14, p. 100522, Mar. 2026, doi: 10.1016/j.fraope.2026.100522.
- [11] A. R. Hermanto, A. Aziz, and S. Sudianto, "Comparison of MobileNetV2 and ResNet50 Architectures for Classification of Date Types," *JUSTIN (Journal of Information Systems and Technology)*, vol. 12, no. 4, pp. 630–637, Nov. 2024, doi: 10.26418/justin.v12i4.80358.
- [12] A. Siti, "Comparison of the Performance of Mobilenetv2 and Resnet50 Architecture in the Classification of Herbal Plants Based on Leaf Image," Jun. 2025.
- [13] Y. Gulzar, "Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique," *Sustainability 2023, Vol. 15, Page 1906*, vol. 15, no. 3, p. 1906, Jan. 2023, doi: 10.3390/su15031906.
- [14] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.
- [15] "Building powerful image classification models using very little data." Accessed: Feb. 10, 2026. [Online]. Available: <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html>.
- [16] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Scientific Reports 2022 12:1*, vol. 12, no. 1, pp. 5979–, Apr. 2022, doi: 10.1038/s41598-022-09954-8.