Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.4, August 2025, pp. 1778~1784
e-ISSN: 2548-6861

Implementation of Convolutional Neural Network in Image-Based Waste Classification

Adila Qurrota A'yun [1]*, Suhartono [2]*, Tri Mukti Lestari [3]*
* Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang
adilahasna1909@gmail.com [1], suhartono@ti.uin-malang.ac.id [2], trimuktilestari@ti.uin-malang.ac.id [3]

Article Info

Article History:

Received 2025-06-15
Revised 2025-07-20
Accepted 2025-07-30

Keyword:

Convolutional Neural Network,
Image Classification,
Waste.

ABSTRACT

The increasingly complex issue of waste management, particularly in the sorting process, demands efficient and accurate technology-based solution. This study aims to implement the Convolutional Neural Network (CNN) method for image-based waste classification, focusing on two classes paper and plastic. The dataset used consists of 2000 images, with an 80% proportion for training and 20% for testing. This study tested four scenarios combining image augmentation and classification methods, namely threshold and one-hot encoding, and evaluated model performance using accuracy, precision, recall, and F1-score metrics. The best results were obtained in the scenario using image augmentation with the one-hot encoding classification method, with an accuracy of 89%, precision of 88.5%, recall of 89%, and F1-score of 88.5%. These findings indicate that implementation of CNN can enhance the effectiveness of image-based waste classification and support recycling efforts through a smarter and more automated sorting system.

I. Introduction

Waste management is a complex environmental issue that is becoming increasingly pressing as the population and human activity grow [1]. According to a report by the World Bank Group, global waste production reached 2.24 billion tons in 2020, with projections indicating an increase to 3.88 billion tons by 2050 [2]. Based on data from the Ministry of Environment and Forestry, the total waste generated in Indonesia reached around 38.79 million tons. However, only 48.45% of this amount was managed optimally [3]. Meanwhile, more than 60% of waste was directly disposed of in landfills without undergoing an adequate sorting process [4].

One of the major challenges in waste management lies in the type of waste that is difficult to decompose, such as plastic. According to a report from Sustainable Waste Indonesia, the amount of plastic waste in Indonesia is estimated to reach 3.2 million tons per year, or approximately 5% of the total waste generated [5]. Plastic waste is a type of inorganic waste that has the potential to be recycled, but it often mixes with other types of waste, making the recycling process difficult. Manual waste sorting by humans requires significant time and effort and carries a high risk of error. Therefore, a technology-based solution is needed to assist in the automatic, fast, and accurate classification of waste.

With the development of technology, artificial intelligence-based approaches such as deep learning have begun to be widely applied in image classification problems. Convolutional Neural Network (CNN) is among the most frequently applied deep learning architectures in image processing due to its strong ability to extract visual features [6]. CNN has the advantage of being able to automatically extract features from images without the need for complex pre-processing [7]. CNN works by utilizing convolution, pooling, and fully connected layers to recognize patterns or distinctive features in images, making it highly suitable for image-based classification tasks [8].

Various studies have proven the effectiveness of the CNN method in image-based waste classification. Dacipta & Putra [9] used the VGGNet architecture to classify waste, including batteries, clothing, e-waste, glass, light bulbs, metal, organic waste, paper, and plastic, achieving an accuracy of 64.45%. Nonetheless, this study faces some limitations, including the absence of data augmentation techniques to increase image diversity and an evaluation that solely focuses on accuracy, without incorporating other performance metrics such as precision, recall, and F1-score. These limitations indicate a research gap that needs to be filled, so this study was conducted to address these shortcomings through the application of image augmentation techniques, the use of more comprehensive evaluation metrics, and a focus on the classification of two types of non-biodegradable waste, namely paper and plastic, which have rarely been discussed in previous studies.

In another study, Yujie He et al. [10] modified the AlexNet architecture for waste classification, including cardboard, glass, metal, paper, plastic, and general waste, and reached an accuracy rate of 79.94%. Additionally, Sutanty & Kusuma Astuti [11] implement the VGG16 architecture to brown glass, cardboard, green glass, metal, paper, plastic, white glass, batteries, masks, and organic waste, achieving an accuracy of 84.62%. Meanwhile, Kurniawan et al. [12] used the Xception architecture for the classification of inorganic waste, namely cardboard, glass, metal, paper, plastic, and waste, achieving an accuracy of 87.81%. These findings indicate that CNN has great potential in supporting automatic waste classification systems, although further optimization is still needed.

Based on the above background, the objective of this research is to apply and assess the performance of the CNN approach in waste image classification tasks, specifically inorganic waste such as paper and plastic. In this study, image augmentation techniques and two classification approaches the threshold method and one-hot encoding were implement to determine their impact on model performance. Implemnet image augmentation strategies alongside two classification approaches is anticipated to offer a more in-depth insight into the overall effectiveness of the implemented methods.

This study not only contributes to waste management systems but also provides a comparative analysis of performance across various testing scenarios. The results of this study are expected to offer an innovative solution to simplify the waste sorting process, thereby supporting recycling efforts and more effective waste management.

II. METHODS

This research was conducted by implement the Convolutional Neural Network (CNN) method to classify waste images into two classes, namely paper and plastic. This method was chosen because CNN has superior capabilities in extracting image features and producing accurate classifications. The research flow generally consists of four stages, namely dataset collection, data preprocessing, CNN model creation, and model performance evaluation.

The research process began with the collection of waste image datasets from online sources, followed by the preprocessing stage to prepare the data for use in model training. Next, the CNN model was built and trained to recognize patterns in images, then tested using several scenarios. The last phase involved evaluating the model's performance using a confusion matrix. Research flow of image-based waste classification using CNN in Figure 1.

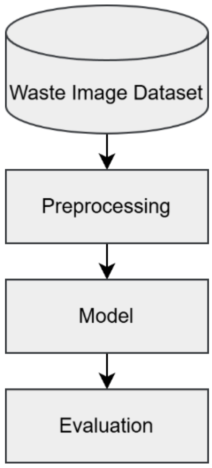

Figure 1. Research Flow of Image-Based Waste Classification Using CNN

A. Dataset

The dataset used comes from the Kaggle platform, namely "Real Waste Image Classification" with an image resolution of 524×524 pixels and "Trash Images" with an image resolution of 384×512 pixels. The data consists of 2000 JPG images divided into two main categories paper and plastic, each containing 1000 images. The dataset was then split into 80% for training and 20% for testing. Sample images of paper and plastic waste from the dataset in Figure 2.

Figure 2. Sample Images of Paper and Plastic Waste from the Dataset

B. Preprocessing

The preprocessing stage is carried out to prepare images for use in model training, while also improving classification efficiency and accuracy. In this study, preprocessing was carried out through the stages of resizing, normalizing, and augmenting images.

Resizing involves changing the dimensions of an image to a specific size so that all images in the dataset are consistent in size. In this study, the image size was changed to 128×128 pixels. This size was chosen based on the findings of

Mishkin et al. [13] which showed that this size is effective in retaining important features in the image while allowing model training to run faster without compromising accuracy.

Normalization aims to change the pixel value scale from a range of 0-255 to a range of 0-1. This step is crucial to support the model's learning efficiency and to speed up the convergence of the loss function during training. The normalization formula is given in equation 1.

$$normalization\_value \frac{pixel\_value}{255} \quad (1)$$

Image augmentation is used to add variety to the training data and improve the model's ability to generalize to new data. This augmentation is especially important to reduce the risk of overfitting, a situation where the model becomes overly tailored to the training data and struggles to detect patterns in unseen testing data. Some augmentation techniques applied include zoom, which is used to enlarge or reduce parts of an image within a range of 20% of the original size so that the model can recognize objects in various sizes. Rotate is used to rotate the image within a range of 90° to help the model understand the orientation of objects that vary. Horizontal flip is used to flip the image horizontally to add data variation. The output of image augmentation techniques in Figure 3.

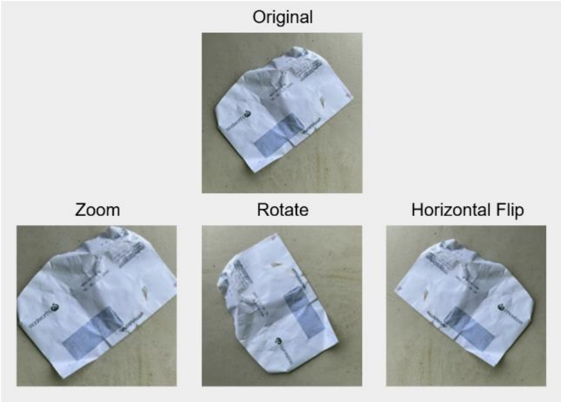

Figure 3. Output of Image Augmentation Techniques

C. Model

This study uses a Convolutional Neural Network (CNN) architecture to extract features from images and perform classification. The CNN architecture used refers to the research by Al-Mamun et al. [14] with a structure consisting of an input, 2 convolutional layers, 2 max pooling layers, a fully connected layer, and an output. The CNN architecture used in this study in Figure 4.

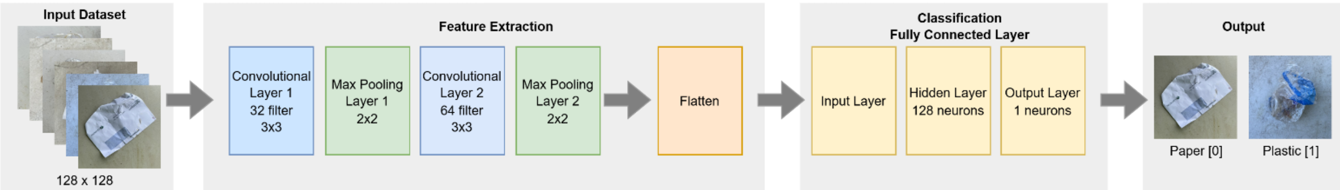

Figure 4. Convolutional Neural Network Architecture Used in This Study

The first layer used is a convolutional layer to extract local features such as edges, patterns, and textures from images. The convolution process is performed using a 3×3 kernel, with a stride of 1 and no padding. The convolution process is calculated in equation 2.

$$FM_{(a,b)} = \left(\sum_{h=0}^{k_h-1} \sum_{w=0}^{k_w-1} K_{(h,w)} \times X_{(a+h,b+w)} \right) \quad (2)$$

After convolution, the ReLU (Rectified Linear Unit) activation function is used to add non-linearity elements. This activation function is expressed in equation 3.

$$f(x) = \max(0, x) \quad (3)$$

Next, a max pooling layer is used to decrease the dimensionality of the feature map and streamline the data representation. Pooling is performed using a 2×2 window and a stride of 2. The pooling process is calculated in equation 4.

$$fP(a, b) = \max \{X(a + h, b + w)\} \quad (4)$$

Before entering the fully connected layer, the output from the pooling layer is first flattened into a one-dimensional vector. In the fully connected layer, the output from each neuron is calculated using equation 5.

$$yt = \sum_{s=1}^{S} w_{t,s} \times h_s + b_1 \quad (5)$$

The next yt output is then processed using the ReLU activation function to add non-linearity using equation 6.

$$h_t = \max(0, yt) \quad (6)$$

The output from the hidden layer is then passed to the output layer, which uses a single neuron with a sigmoid activation function. The output value before activation is calculated using equation 7.

$$z_o = \sum_{t=1}^{T} w(o,t) \times h_t + b_2 \tag{7}$$

Then converted to probability using the sigmoid function in equation 8.

$$\hat{y} = \sigma(z_o) = \frac{1}{1 + e^{-z_o}} \tag{8}$$

The classification result is determined based on a threshold value of 0.5. The selection of this value refers to Brownlee [15] explanation that in the sigmoid activation function, the value 0.5 is the midpoint that divides the output range [0,1] symmetrically, so it is used as the boundary for binary classification. Therefore, if $\hat{y} > 0.5$, the image is classified as plastic (label 1), while if $\hat{y} \leq 0.5$, the image is classified as paper (label 0).

Then, the model was trained using parameters set at a constant value to maintain consistency in performance evaluation. The values of each parameter in Table 1.

TABLE I
VALUE PARAMETER

Parameter	Value Parameter
Batch Size	32
Dropout	0.5
Activation Function	ReLu, Sigmoid, Softmax
Optimizer	Adam
Learning Rate	0.001
Epoch	25

In addition to designing the model architecture and setting constant training parameters, this study also developed four testing scenarios to evaluate the effect of image augmentation techniques and classification methods on model performance. These scenarios are distinguished based on the combination of image augmentation and the threshold classification method using a threshold of 0.5 to assign the predicted class label, as well as the one-hot encoding classification method, which compares the highest predicted value in the output layer to determine the predicted class. The threshold classification uses a sigmoid activation function in the output layer, while the one-hot encoding method uses a softmax activation function. The combinations of test scenarios in Table 2.

TABLE II
COMBINATIONS OF TEST SCENARIOS

Combination	Combinations of Test Scenarios
1A	Without image augmentation using the threshold classification method
1B	Without image augmentation using the one-hot encoding classification method
2A	Using image augmentation with the threshold classification method
2B	Using image augmentation with the one-hot encoding classification method

D. Evaluation

A model's performance can be assesse through a number of metrics such as accuracy, precision, recall, and F1-score, which are obtained based on confusion matrix analysis. The confusion matrix includes several important values, namely:

True Positive (TP)	:	the number of plastic images that are correctly classified as plastic.
True Negative (TN)	:	the number of paper images that are correctly classified as paper.
False Positive (FP)	:	the number of paper images incorrectly classified as plastic.
False Negative (FN)	:	the number of plastic images incorrectly classified as paper.

From these values, the accuracy level, which indicates how well the model performs classification, is calculated using equation 9.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Next, precision indicates the accuracy of the model in predicting positive classes, which is the percentage of correct positive predictions compared to all positive predictions generated by the model, calculated using equation 10.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

In addition, recall indicates how many positive events were successfully detected by the model compared to all positive events that actually existed, calculated using equation 11.

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

The F1-score provides a balance between precision and recall, calculated using equation 12.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

III. RESULTS AND DISCUSSIONS

This research examines the effectiveness of the CNN model in classifying waste images into two classes, namely paper and plastic, through four testing scenarios. Each scenario is a combination of two approaches the use of image augmentation and the threshold one-hot encoding classification method. The model evaluation was carried out using a simple 80/20 train-test split, without applying cross-validation techniques. The results obtained include training loss, training accuracy, confusion matrix, and evaluation metrics that include precision, recall, and F1-score.

A. *Test Scenario 1A*

Scenario 1A was conducted without image augmentation using the threshold classification method. The training phase resulted in a loss value of 0.3545 and an accuracy of 85.59%, as shown in Figure 5.

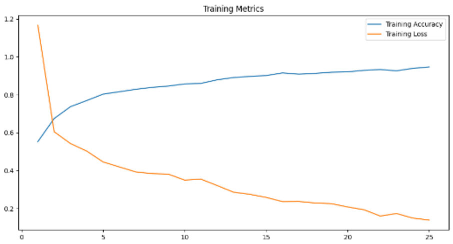

Figure 5. Training Loss and Accuracy Graph for Scenario 1A

The classification results are visualized through a confusion matrix in Figure 6, where the model successfully predicted 162 paper samples and 174 plastic samples correctly, and made errors in 42 paper and 22 plastic samples.

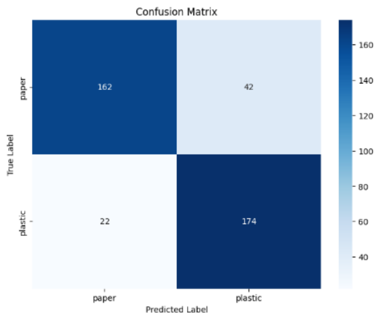

Figure 6. Confusion Matrix for Scenario 1A

During testing, the model achieved a testing accuracy of 84%. Evaluation metrics precision, recall, and F1-score Table 3.

TABLE III
METRICS EVALUATION TEST SCENARIO 1A

Class	**Precision**	**Recall**	**F1-score**
Paper	88%	79%	84%
Plastic	81%	89%	84%

Without augmentation, the model learns from data that is limited in variety. This is reflected in the difference in precision and recall between classes. The model appears to be more confident in classifying images as plastic, which results in high recall for plastic but low precision, which is consistent with the error pattern in the confusion matrix.

B. *Test Scenario 1B*

Scenario 1B was conducted without image augmentation, but classification was performed using the one-hot encoding method. The training process resulted in a training loss of 0.3582 and a training accuracy of 85.97%, as shown in Figure 7.

Figure 7. Training Loss and Accuracy Graph for Scenario 1B

The classification results are visualized through a confusion matrix in Figure 8. The model successfully predicted 163 papers and 180 plastics correctly, while 41 papers and 16 plastics were classified incorrectly.

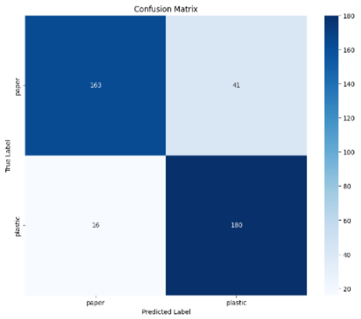

Figure 8. Confusion Matrix for Scenario 1B

During testing, the model achieved a testing accuracy of 86%. Evaluation metrics precision, recall, and F1-score Table 4.

TABLE IV
METRICS EVALUATION TEST SCENARIO 1B

Class	**Precision**	**Recall**	**F1-score**
Paper	91%	80%	85%
Plastic	81%	92%	86%

One-hot encoding helps reduce ambiguity in the classification process by explicitly representing classes as binary vectors. This results in an increase in recall and F1-score values compared to scenario 1A. Even without image augmentation, these results show that one-hot encoding contributes positively to label representation and model learning.

C. *Test Scenario 2A*

Scenario 2A applies image augmentation and uses the threshold classification method. The training results produced a training loss of 0.3694 and a training accuracy of 85.54%, as shown in Figure 9.

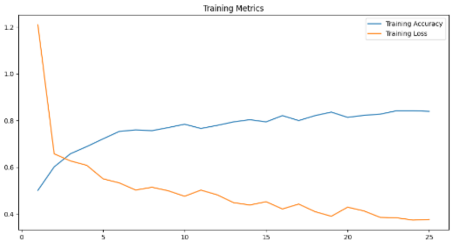

Figure 9. Training Loss and Accuracy Graph for Scenario 2A

The confusion matrix presented in Figure 10 demonstrates that the model successfully predicted 167 papers and 172 plastics correctly, with errors in 37 papers and 24 plastics.

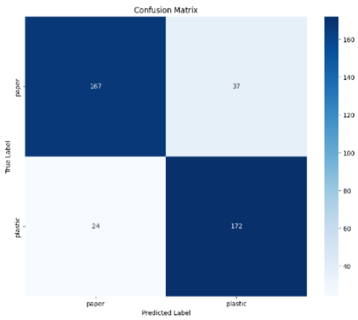

Figure 10. Confusion Matrix for Scenario 2A

During testing, the model achieved a testing accuracy of 85%. Evaluation metrics precision, recall, and F1-score Table 5.

TABLE V
METRICS EVALUATION TEST SCENARIO 2A

Class	Precision	Recall	F1-score
Paper	87%	82%	85%
Plastic	82%	88%	85%

By applying image augmentation, the model learns from more varied data, which has been proven to improve recall and F1-score. However, because classification still uses thresholds, the model still has limitations in recognizing images with ambiguous characteristics or those that are on the threshold of classification.

D. Test Scenario 2B

Scenario 2B combines image augmentation with the one-hot encoding classification method. The model produces the lowest training loss of 0.3255 and the highest training accuracy of 89.29%, as shown in Figure 11.

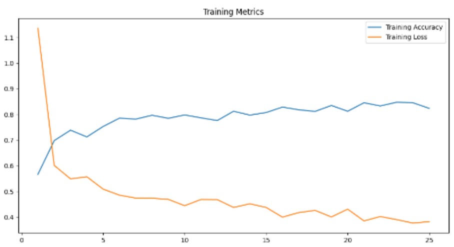

Figure 11. Training Loss and Accuracy Graph for Scenario 2B

As seen in Figure 12, the confusion matrix reveals that the model successfully predicted 183 papers and 172 plastics, with errors only in 21 papers and 24 plastics.

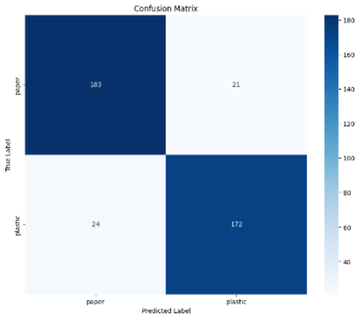

Figure 12. Confusion Matrix for Scenario 2B

During testing, the model achieved a testing accuracy of 89%. Evaluation metrics precision, recall, and F1-score Table 6.

TABLE VI
METRICS EVALUATION TEST SCENARIO 2B

Class	Precision	Recall	F1-score
Paper	88%	90%	89%
Plastic	89%	88%	88%

Scenario 2B shows the best results. Image augmentation enriches the diversity of image features and helps the model recognize more visual variations. Meanwhile, the one-hot encoding classification method provides a more distinct label representation and reduces prediction ambiguity. This combination allows the model to produce more balanced, accurate, and consistent classifications, as reflected in the high and balanced F1-score values for both classes.

To determine the differences in performance between each test scenario, Table 7 presents a comparison of accuracy, precision, recall, and F1-score values.

TABLE VII
COMPARISON OF ACCURACY AND EVALUATION METRICS BETWEEN SCENARIOS

Scenarios	Accuracy	Precision	Recall	F1-score
1A	84%	84.5%	84%	84%
1B	86%	86%	86%	85.5%
2A	85%	84.5%	85%	85%
2B	89%	88.5%	89%	88.5%

Based on the table above, scenario 2B consistently produces the best values across all evaluation metrics, as supported by the training graph and confusion matrix. Image augmentation improves the model's generalization ability by adding variation to the training data, so that the model does not just memorize patterns but actually learns to recognize the characteristics of each class. Meanwhile, the one-hot encoding classification method helps avoid bias from fixed threshold values, which are often too coarse binary boundaries for complex image classification cases. The combination of both allows the model to adapt better to varied images. To evaluate the position and contribution of this study, Table 8 presents a comparison of the classification accuracy of several previous studies.

TABLE VIII
COMPARISON OF CLASSIFICATION PERFORMANCE WITH PREVIOUS STUDIES

Study	Model	Accuracy
Dacipta & Putra [9]	VGGNet	64.45%
Yujie He et al. [10]	Modified AlexNet	79.94%
Sutanty & Astuti [11]	VGG16	84.62%
Kurniawan et al. [12]	Xception	87.81%
This Study	Custom CNN	89%

Based on Table VIII, the model proposed in this study shows higher accuracy compared to several previous studies. This indicates that a custom-designed, simple CNN architecture with the right augmentation and classification strategies has the potential to be more computationally efficient and capable of producing competitive results, even without using a pre-trained model.

IV. CONCLUSION

The findings of this study confirm that implement a Convolutional Neural Network (CNN) is effective in classifying waste images into two classes, namely paper and plastic. Through four testing scenarios combining image augmentation and threshold one-hot encoding classification methods, the results show that the combination of image augmentation and one-hot encoding classification in scenario 2B yields the best performance with an accuracy of 89%. Image augmentation was proven to enhance the model's generalization to data variations, while the one-hot encoding classification method provided a more explicit and stable label representation. The evaluation based on accuracy, precision, recall, and F1-score metrics indicates that combining both approaches led to the most accurate and balanced classification outcomes. Therefore, it can be concluded that CNN is an effective method for application in image-based waste classification systems.

Based on the results obtained and considering a number of limitations in this study, several suggestions can be made for further research development, namely adding image augmentation variations, such as shear, crop, and vertical flip, to enhance the model's capability in recognizing diverse types of images. Additionally, a more complex CNN architecture could be implemented by adding convolutional layers or other architectures such as VGG16 and ResNet to improve the model's accuracy and generalization. Further research is also recommended to test the model on new data (unseen data) in order to evaluate the model's performance more comprehensively in real-world conditions.

REFERENCES

[1] D. Deslanda, P. Mukti, M. Kaisar, N. D. Lestari, and S. Salma, "Sosialisasi Pengolahan Sampah Organik dan Anorganik Menjadi Produk Kreatif," *Madani J. Ilm. Multidisipline*, vol. 1, no. 11, pp. 745–749, 2023.

[2] World Bank Group, "Solid Waste Management," World Bank Group. Accessed: Jun. 13, 2025. [Online]. Available: https://www.worldbank.org/en/topic/urbandevelopment/brief/solid-waste-management

[3] Kementerian Lingkungan Hidup dan Kehutanan, "Capaian Kinerja Pengelolaan Sampah Indonesia," Kementerian Lingkungan Hidup dan Kehutanan. Accessed: Jun. 13, 2025. [Online]. Available: https://sipsn.menlhk.go.id/sipsn/

[4] H. K. N. Sumartiningtyas, "Indonesia Hasilkan 64 Juta Ton Sampah, Bisakah Kapasitas Pengelolaan Tercapai Tahun 2025?," Kompas.com. Accessed: Jun. 13, 2025. [Online]. Available: https://www.kompas.com/sains/read/2020/12/18/070200023/indonesia-hasilkan-64-juta-ton-sampah-bisakah-kapasitas-pengelolaan?page=all

[5] D. Waluyo, "Masalah Kita Adalah Sampah," Indonesia.Go.Id. Accessed: Jun. 13, 2025. [Online]. Available: https://www.indonesia.go.id/kategori/editorial/7714/masalah-kita-adalah-sampah.

[6] J. Yan, "Application of CNN in computer vision," *Appl. Comput. Eng.*, vol. 30, no. 1, pp. 104–110, 2024, doi: 10.54254/2755-2721/30/20230081.

[7] O. D. S. Sunanto and P. H. Utomo, "Implementasi Deep Learning Dengan Convolutional Neural Network Untuk Klasifikasi Gambar Sampah Organik Dan Anorganik," *Pattimura Proceeding Conf. Sci. Technol.*, vol. 1, no. 2, pp. 335–340, 2022, [Online]. Available: https://jurnal.unej.ac.id/index.php/prosiding/article/view/33527

[8] A. Peryanto, A. Yudhana, and R. Umar, "Klasifikasi Citra Menggunakan Convolutional Neural Network dan K Fold Cross Validation," *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 45–51, 2020, doi: 10.30871/jaic.v4i1.2017.

[9] P. N. Dacipta and R. E. Putra, "Sistem Klasifikasi Limbah Menggunakan Metode Convolutional Neural Network (CNN) Pada Webservice Berbasis Framework Flask," *J. Informatics Comput. Sci.*, vol. 3, no. 04, pp. 394–402, 2022, doi: 10.26740/jinacs.v3n04.p394-402.

[10] Yujie He, Qinyue Gu, and Maguo Shi, "Trash Classification Using Convolutional Neural Networks Project Category: Computer Vision," *CS230 Deep Learn.*, 2020.

[11] E. Sutanty and D. Kusuma Astuti, "DECODE: Jurnal Pendidikan Teknologi Informasi Penerapan Model Arsitektur VGG16 Untuk Klasifikasi Jenis Sampah," vol. 3, no. 2, pp. 407–419, 2023.

[12] R. Kurniawan, P. B. Wintoro, Y. Mulyani, and M. Komarudin, "Implementasi Arsitektur Xception Pada Model Machine Learning Klasifikasi Sampah Anorganik," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 2, pp. 233–236, 2023, doi: 10.23960/jitet.v11i2.3034.

[13] D. Mishkin, N. Sergievskiy, and J. Matas, "Systematic evaluation of CNN advances on the ImageNet," 2016, doi: 10.1016/j.cviu.2017.05.007.

[14] A.-M. Al-Mamun, R. Hossain, M. M. M. Sharmin, E. Kabir, and M. A. Iqbal, "Garbage classification using convolutional neural networks (CNNs)," *Mater. Sci. Eng. Int. J.*, vol. 7, no. 3, pp. 140–144, 2023, doi: 10.15406/mseij.2023.07.00217.

[15] J. Brownlee, "A Gentle Introduction to Threshold-Moving for Imbalanced Classification," Machine Learning Mastery. Accessed: Jun. 13, 2025. [Online]. Available: https://machinelearningmastery.com/threshold-moving-for-imbalanced-classification/