Implementation of Transfer Learning Using ResNet-18 for Image-Based Garbage Classification

Dika Nur Kholish[1]
[1] Departement of Electrical Engineering, Faculty of Engineering, Universitas Singaperbangsa Karawang, Karawang, Indonesia
[1] email: 2310631160052@student.unsika.ac.id

[*submitted*: 16-02-2026 | *review*: 11-04-2026 | *published*: 30-05-2026]

ABSTRACT: *Effective waste management requires accurate and efficient classification systems to support recycling and environmental sustainability efforts. This study presents the implementation of a transfer learning approach using the ResNet-18 convolutional neural network for image-based garbage classification. A publicly available garbage classification dataset consisting of multiple waste categories was utilized. The dataset was divided into training and validation sets with an 80:20 ratio. Data preprocessing included image resizing, normalization, and augmentation techniques such as random horizontal flipping and rotation to improve generalization. The pretrained ResNet-18 model was adapted by freezing convolutional layers and replacing the fully connected layer according to the number of classes. The model was trained using the Adam optimizer and cross-entropy loss function. Experimental results demonstrate that the proposed approach achieved a validation accuracy of approximately 93%, indicating strong classification performance. Analysis of the confusion matrix reveals that most classes were correctly identified, with minor misclassifications occurring between visually similar categories. These findings confirm that transfer learning with ResNet-18 provides an effective and computationally efficient solution for garbage image classification tasks.*

KEYWORDS: *Convolutional neural networks; Transfer learning; Garbage classification; Image classification; ResNet-18;*

I. INTRODUCTION

The increasing volume of municipal solid waste has become a critical environmental challenge, requiring efficient waste management and recycling strategies. Accurate waste classification plays an essential role in improving recycling processes; however, manual sorting methods are inefficient, time-consuming, and susceptible to human error [1]. As a result, automated waste classification systems based on image analysis have gained significant research attention.

Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have demonstrated superior performance in image classification tasks due to their ability to automatically learn discriminative visual features [2]. Nevertheless, training deep CNN models from scratch demands large labeled datasets and substantial computational resources, which are often unavailable in practical applications. Transfer learning addresses this issue by leveraging pretrained models trained on large-scale datasets such as ImageNet.

Residual Networks (ResNet), introduced by He *et al.* [3], employ residual connections to facilitate efficient training while maintaining high classification accuracy. Among the available architectures, ResNet-18 offers a favorable balance between model complexity and performance, making it suitable for applied image classification tasks. Therefore, this study implements a transfer learning approach using a pretrained ResNet-18 model for image-based garbage classification and evaluates its effectiveness using standard performance metrics.

II. OVERVIEW

A. IMAGE CLASSIFICATION

Image classification is a fundamental task in computer vision that aims to assign a predefined label to an image based on its visual content. The objective is to identify discriminative patterns such as color, texture, and shape that distinguish one object category from another. In environmental applications, image classification plays an important role in enabling automated waste sorting systems, reducing reliance on manual segregation processes and improving operational efficiency [1].

Traditional image classification approaches relied on handcrafted feature extraction methods combined with classical machine learning algorithms. However, these approaches often struggle to generalize across diverse and complex datasets. The breakthrough in large-scale image recognition was achieved with the introduction of deep Convolutional Neural Networks (CNNs), particularly after the success of AlexNet on the ImageNet benchmark [2], [4]. These models demonstrated that hierarchical feature learning directly from raw image data significantly improves classification accuracy.

© the Authors (2026)
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

B. CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional Neural Networks (CNNs) are a class of deep learning architectures specifically designed for processing grid-structured data such as images. CNNs have become the dominant approach in image classification due to their capability to automatically learn hierarchical feature representations through multiple convolutional layers [2]. Unlike traditional machine learning methods that rely on handcrafted features, CNNs perform end-to-end learning, extracting low-level features such as edges and textures in early layers and more abstract semantic features in deeper layers.

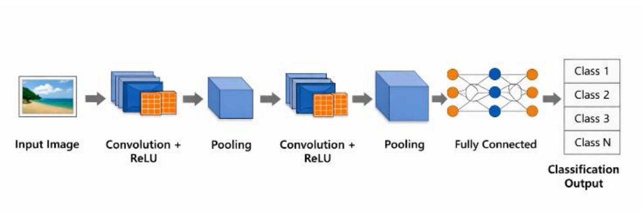

Fig 1. Basic Architecture of a Convolutional Neural Network (CNN) for Image Classification

A typical CNN architecture consists of convolutional layers, pooling layers, and fully connected layers. Convolutional layers apply learnable filters to capture spatial patterns, while pooling layers reduce spatial dimensions and improve translation invariance. The fully connected layers then perform classification based on the learned feature maps. The effectiveness of CNNs was notably demonstrated in large-scale image recognition tasks, particularly in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [2], [4].

However, increasing network depth may lead to training difficulties such as vanishing gradients. Residual learning, introduced in ResNet architectures, addresses this problem by incorporating shortcut connections that enable more stable and efficient optimization [3]. This advancement significantly improved the performance of deep CNN models and facilitated their application in various computer vision tasks, including object recognition and environmental image classification.

C. TRANSFER LEARNING

Transfer learning is a machine learning paradigm in which knowledge gained from solving one problem is applied to a different but related task [5]. In the context of deep learning, transfer learning typically involves utilizing a model pretrained on a large-scale dataset, such as ImageNet, and adapting it to a target task with a smaller dataset. This approach significantly reduces the need for extensive labeled data and computational resources while maintaining high performance.

Deep CNN models trained on ImageNet learn generic low-level and mid-level visual features that are transferable across various image classification problems [2], [4]. By reusing pretrained convolutional layers and retraining only the final classification layer, the model can effectively adapt to new categories with limited training data. This strategy is particularly beneficial in practical applications where dataset size and computational capacity are constrained.

In this study, transfer learning is implemented by employing a pretrained ResNet-18 model and replacing its fully connected layer to match the number of garbage classes. The convolutional layers are frozen during training to preserve the learned feature representations, while the final classification layer is optimized for the specific garbage classification task. This approach provides a balance between computational efficiency and classification accuracy.

D. RESNET-18 ARCHITECTURE

Residual Networks (ResNet) were introduced to address the degradation problem in deep neural networks, where increasing network depth leads to higher training error due to optimization difficulties [3]. The core innovation of ResNet lies in the use of residual connections, also known as shortcut connections, which allow the input of a layer to bypass one or more intermediate layers and be added directly to the output. This mechanism facilitates gradient flow during backpropagation and mitigates the vanishing gradient problem.

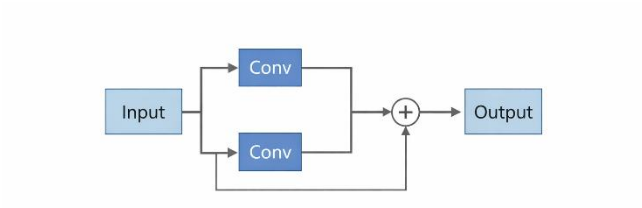

Fig 2. Ilustration of a Residual Block with Shortcut Connection

ResNet-18 is one of the lightweight variants of the ResNet family, consisting of 18 convolutional layers organized into residual blocks. Each residual block typically contains two convolutional layers with batch normalization and ReLU activation, followed by an identity shortcut connection. Compared to deeper variants such as ResNet-50 or ResNet-101, ResNet-18 has lower computational complexity while maintaining competitive performance in image classification tasks.

Due to its balanced architecture and reduced parameter size, ResNet-18 is well suited for practical applications where computational efficiency is an important consideration. In this study, a pretrained ResNet-18 model is utilized as the feature extractor, and its final fully connected layer is replaced to accommodate the number of garbage classes. This

© the Authors (2026)
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

configuration enables effective feature transfer while maintaining a relatively lightweight training process.

III. METHODOLOGY

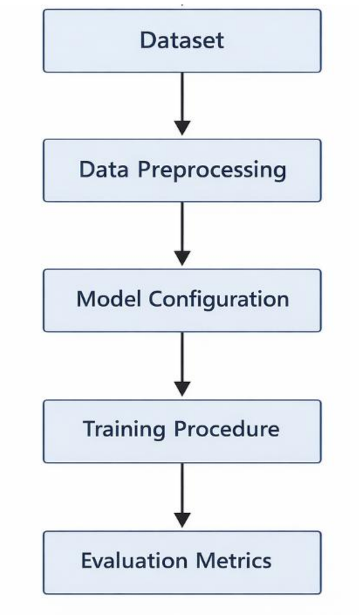

Fig 3. Research Workflow

A. DATASET

The dataset consists of multiple categories of waste images representing common recyclable and non-recyclable materials. Each class contains images captured under varying lighting conditions, orientations, and backgrounds, providing diversity in visual characteristics.

The dataset is divided into training and validation sets using an 80:20 split ratio. The training set is used to optimize the model parameters, while the validation set is employed to evaluate generalization performance during training.

B. DATA PREPROCESSING

All input images are resized to 224 × 224 pixels to match the input dimension required by the ResNet-18 architecture. To improve model generalization and reduce overfitting, data augmentation techniques are applied to the training dataset, including random horizontal flipping and random rotation within a limited angle range.

After augmentation, images are converted into tensor format and normalized using the standard ImageNet mean and standard deviation values. Normalization ensures that the input distribution is consistent with the pretrained ResNet-18 model.

C. MODEL CONFIGURATION

A pretrained ResNet-18 model is employed as the base architecture. The convolutional layers are initialized with weights trained on the ImageNet dataset. To implement transfer learning, all convolutional layers are frozen to preserve previously learned feature representations.

The original fully connected layer of ResNet-18 is replaced with a new linear layer corresponding to the number of garbage categories in the dataset. Only this final classification layer is trained during the optimization process.

D. TRAINING PROCEDURE

Model training is conducted for five epochs using the Adam optimizer with a learning rate of 0.001. The cross-entropy loss function is employed to measure classification error. A batch size of 16 is used during training.

During each epoch, the model parameters of the fully connected layer are updated through backpropagation, while the frozen convolutional layers remain unchanged. After each epoch, validation loss and accuracy are computed to monitor model performance and convergence behavior.

E. EVALUATION METRICS

Model performance is evaluated using validation accuracy and loss values. Accuracy measures the proportion of correctly classified samples relative to the total number of validation samples.

In addition, a confusion matrix is generated to analyze class-wise prediction performance and identify misclassification patterns. This analysis provides insight into similarities between visually related categories and potential classification challenges.

IV. RESULTS AND DISCUSSION

A. TRAINING AND VALIDATION PERFORMANCE

Fig 4. Training vs Validation Loss of the Model

© the Authors (2026)
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

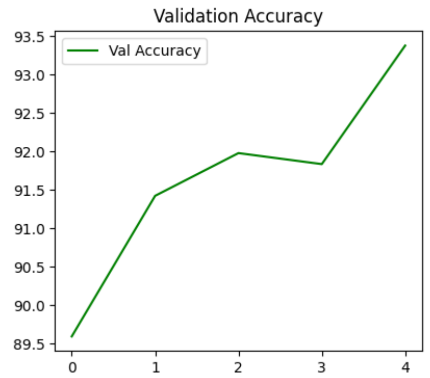

Fig 5. Validation Accuracy

The training dynamics of the proposed transfer learning model are illustrated in Figure 4 and Figure 5. A consistent reduction in training loss is observed, decreasing from approximately 0.67 in the first epoch to 0.34 in the fifth epoch. This downward trend indicates that the model successfully minimizes the optimization objective and progressively learns discriminative features from the training data.

More importantly, the validation loss follows a similar decreasing pattern, declining from approximately 0.35 to 0.22. The relatively small gap between training and validation loss throughout the epochs suggests that the model maintains good generalization capability. The absence of divergence between the two curves indicates that overfitting does not occur within the observed training duration. This behavior confirms that freezing the convolutional layers while fine-tuning only the fully connected layer is sufficient for adapting pretrained features to the garbage classification task.

The validation accuracy improves steadily from 89.6% in the first epoch to 93.4% in the final epoch. The rapid performance improvement within only five epochs demonstrates the effectiveness of transfer learning in accelerating convergence. Since the convolutional layers retain pretrained weights learned from large-scale visual data, the model primarily adjusts the final classification layer to align with the new class distribution. This significantly reduces the number of trainable parameters and stabilizes the learning process.

A slight fluctuation in validation accuracy between the third and fourth epochs can be observed; however, this variation is marginal and does not indicate instability. Instead, it reflects normal stochastic behavior during mini-batch optimization.

B. CLASSIFICATION PERFORMANCE ANALYSIS

To obtain a more comprehensive understanding of model behavior, the confusion matrix presented in Figure Z is analyzed in detail. The matrix reveals strong diagonal dominance, indicating that most samples are correctly classified across the majority of waste categories. This confirms that the pretrained ResNet-18 successfully extracts discriminative visual features that generalize well to the validation dataset.

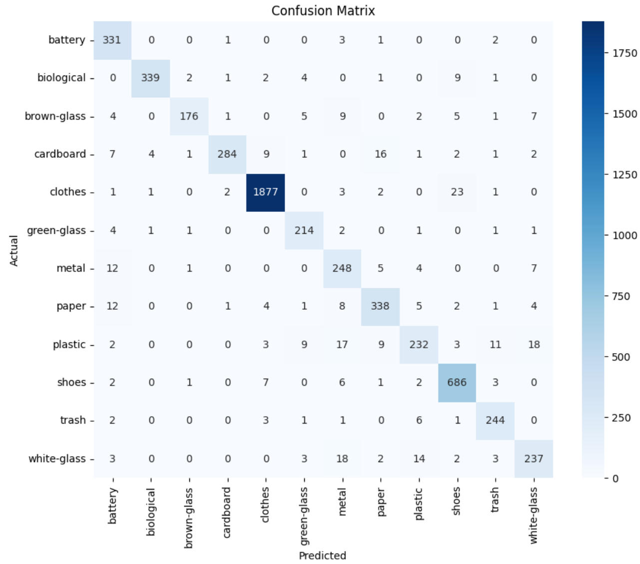

Fig 6. Confusion Matrix from the Model

Several categories exhibit particularly high classification accuracy. For instance, the *clothes* class achieves the highest number of correct predictions, suggesting that its visual characteristics—such as texture patterns and shape irregularities—are highly distinguishable from other waste categories. Similarly, *shoes* and *biological* waste classes demonstrate strong classification consistency, likely due to distinctive structural features and relatively low inter-class visual similarity.

However, closer inspection of the confusion matrix reveals systematic misclassification patterns among visually similar materials. The most notable confusion occurs between *plastic*, *white-glass*, and *metal*. These categories often share reflective surfaces, smooth textures, and comparable geometric shapes, particularly in cases of packaging materials. Such similarities reduce inter-class separability in the feature space, causing overlapping representations in the learned embedding space of the network.

Additionally, minor confusion between *brown-glass* and *green-glass* can be attributed to color variations influenced by lighting conditions. Since pretrained CNN models primarily learn texture and edge-based representations rather than absolute color invariance, subtle illumination differences may reduce classification precision in these categories.

From a representation learning perspective, the observed misclassifications suggest that while the pretrained convolutional layers effectively capture general visual structures, fine-grained material differentiation may require additional feature

© the Authors (2026)
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

refinement. Fine-tuning deeper convolutional layers or incorporating attention mechanisms could potentially improve separability among visually overlapping classes.

Despite these localized errors, the overall confusion remains limited, and class-wise performance remains consistently high. The distribution of misclassifications indicates that errors are not random but structurally explainable, reinforcing the validity of the model's learning behavior. This structured error pattern further confirms that the model is not underfitting but rather constrained by intrinsic visual similarity between certain waste materials.

C. QUALITATIVE PREDICTION EXAMPLE

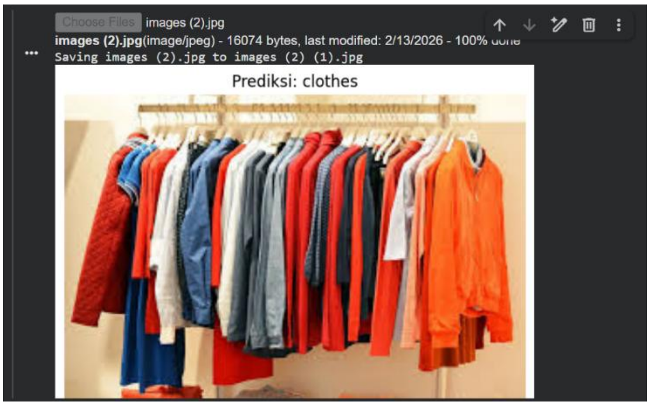

Fig 7. Example of garbage image prediction classified as "clothes" by the proposed ResNet-18 model

To further demonstrate the practical capability of the proposed model, a qualitative prediction example is presented in Figure 7. The input image consists of multiple clothing items displayed on a rack. The model successfully classifies the image as *clothes*, which aligns with the actual category.

This result indicates that the pretrained ResNet-18 model effectively captures texture, shape, and structural features characteristic of clothing materials. The correct classification in this example supports the quantitative findings presented in the previous sections and confirms the robustness of the learned feature representations.

V. CONCLUSION

This study presented the implementation of a transfer learning approach using a pretrained ResNet-18 model for image-based garbage classification. By leveraging pretrained convolutional layers and retraining only the final fully connected layer, the proposed method achieved efficient convergence within five training epochs. Experimental results demonstrated a validation accuracy of approximately 93.4%, accompanied by stable training and validation loss curves, indicating strong generalization capability without significant overfitting.

The confusion matrix analysis revealed high classification performance across most garbage categories, particularly for visually distinctive classes such as clothes and shoes. Misclassification patterns were primarily observed among visually similar materials, including plastic, metal, and glass, reflecting intrinsic feature overlap rather than model instability. These findings confirm that transfer learning with ResNet-18 provides robust feature representations for multi-class waste image classification.

From a practical perspective, the lightweight configuration and reduced training complexity highlight the suitability of the proposed approach for real-world applications, especially in environments with limited computational resources. Although further improvements may be achieved through fine-tuning strategies or architectural comparisons, the results demonstrate that the proposed system offers an effective and computationally efficient solution for automated garbage classification tasks.

Future work may focus on enhancing fine-grained material discrimination, incorporating attention mechanisms, and evaluating the system under more diverse real-world conditions to further strengthen its applicability in intelligent waste management systems.

REFERENCES

[1] A. M. Zaman, "A comprehensive review of solid waste management systems," *Waste Management*, vol. 33, no. 4, pp. 988–1003, 2013, doi: 10.1016/j.wasman.2012.11.007.
[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, 2012, pp. 1097–1105.
[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.
[5] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and*

© the Authors (2026)
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.

AUTHORS BIOGRAPHY AND CONTRIBUTIONS

Dika Nur Kholish is an undergraduate student at Universitas Singaperbangsa Karawang (UNSIKA), Indonesia, with strong interests in technology-driven innovation, Internet of Things (IoT), artificial intelligence, sustainable agriculture, and digital health. His academic and research activities focus on the application of smart technologies to address real-world challenges in agriculture, health, and environmental sustainability. He has actively participated in and received awards from various national and university-level scientific competitions, including scientific writing, engineering innovation, business plan development, and documentary-based innovation projects. His research works cover topics such as IoT-based pest control systems, smart farming 4.0, non-invasive glucose sensing with AI, sustainable agribusiness, and bioplastic-based agricultural innovations.

© the Authors (2026)
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.