

CNN Optimization Based on Transfer Learning through a Combination of Specialized Augmentation and Fine-Tuning for Organic-Inorganic Waste Classification

¹Muhammad Luthfillah Andria, ²Doni Winarso & ³Rizka Hafsari
^{1,2&3} Computer Science, Universitas Muhammadiyah Riau, Indonesia

*¹220402067@student.umri.ac.id

²doniwinarso@umri.ac.id

³rizkahafsari@umri.ac.id

*Corresponding author

ABSTRACT

Waste management in Indonesia still relies heavily on manual sorting between organic and inorganic waste, which is slow, inconsistent, and prone to error. This study develops a Convolutional Neural Network (CNN) classification model for organic and inorganic waste images using a ResNet-50 backbone with transfer learning, optimized through a combination of two-phase fine-tuning and class-specific data augmentation. The dataset, Waste Classification Data from Kaggle (25,077 images), was re-split by stratified random sampling into training (17,553), validation (3,762), and test (3,762) subsets. Augmentation was tailored per class: the Organic class received both geometric and color/channel transformations to simulate decay-related color changes, while the Inorganic class received only geometric transformations to preserve material-specific color as a discriminative feature. Training proceeded in two stages: feature extraction with a fully frozen backbone, followed by gradual fine-tuning of the conv4 block and above using a reduced learning rate. An ablation comparison showed that a baseline model (no fine-tuning, no augmentation) reached only 78.02% accuracy, whereas the proposed model achieved 96.23% accuracy, 96.29% precision, 96.23% recall, and a 96.23% F1-score, an 18.21-percentage-point accuracy gain over the baseline. The improvement was consistent across both classes, most notably raising Inorganic recall by 20.52 points. These results confirm that combining gradual fine-tuning with class-specific augmentation substantially improves the generalization ability of a ResNet-50-based CNN for organic-inorganic waste classification, yielding performance competitive with prior transfer-learning approaches on similar tasks.

Keywords: Data Augmentation, Fine-Tuning, ResNet-50, Transfer Learning, Waste Classification.

INTRODUCTION

Waste management remains a growing global challenge as population and consumption levels increase. Waste is broadly divided into organic waste, which decomposes biologically, and inorganic waste, which resists natural decomposition. Effective management depends on accurate separation between these two categories, since each requires a different downstream process composting for organic waste versus mechanical recycling for inorganic waste. In practice, however, sorting is still largely performed manually, a process vulnerable to inconsistency, fatigue, and misclassification, which limits the

recovery of valuable recyclable material [1], [2]. This has motivated growing interest in automated, image-based sorting solutions built on artificial intelligence [1].

Among image classification approaches, conventional methods such as Support Vector Machine and K-Nearest Neighbor require manual feature engineering, while Random Forest performs sub-optimally on high-dimensional pixel data without prior feature extraction (Alzubaidi et al., 2021). Sequence-oriented architectures such as Recurrent Neural Networks are not designed for spatial image data, and Vision Transformers require very large training sets to reach competitive performance [3], [4]. Convolutional Neural Networks (CNNs), in contrast, learn relevant visual features directly from raw pixels through hierarchical local pattern detection, and have been shown to consistently outperform both conventional machine learning and sequential architectures on image classification tasks [3]. This advantage is particularly relevant to waste identification and recycling, where CNNs have become the dominant and most reliable approach [5], motivating their selection as the core algorithm in this study.

Despite these strengths, CNNs require substantial training data to avoid overfitting and are computationally expensive to train from scratch. A further limitation central to this study is limited generalization when training data does not reflect real-world variability in lighting, camera angle, and background [3]. This is especially relevant for waste imagery, where lighting changes, decay-induced color shifts in organic waste, and heterogeneous backgrounds are common in the field but under-represented in controlled training sets. Color-space augmentation has been shown to broaden the color distribution represented in training data, improving robustness to unanticipated lighting and color variation [6], and high visual variability arising from material type, color, and degradation has been identified as a major obstacle to accurate waste classification [7], underscoring the need for augmentation tailored to this domain.

Prior studies have demonstrated the promise of CNNs for waste image classification. Wu and Lin (2022) compared VGGNet-16 and ResNet-50 for organic-versus-residual waste classification and found ResNet-50 with transfer learning achieved the highest accuracy. Lin et al. (2022) developed a ResNet-based transfer-learning model that achieved strong precision, recall, and F1-scores on a recycling dataset. However, most of these studies applied generic augmentation configurations that were not adapted to the distinct visual characteristics of organic versus inorganic waste, and none examined in depth how fine-tuning interacts with domain-specific augmentation.

Building on this gap, the present study focuses on improving the generalization of a CNN for organic-inorganic waste classification through a training configuration adapted to the visual characteristics of waste imagery. The Waste Classification Data from Kaggle was chosen for its sufficient volume of labeled samples, allowing the study to focus on model optimization rather than data collection. To the best of the author's literature review, this specific combination of training configurations has not previously been applied together to a ResNet-50 backbone on this dataset. Beyond its technical contribution, this study is positioned within Information Systems research as a component of a broader waste-management information system: the resulting model functions as an automated decision-making component that converts raw waste images into actionable classification output, supporting operational decisions such as recycling routing and category-based waste reporting, and connecting to the concept of Green Information Systems.

This study addresses the following research question: to what extent does combining gradual fine-tuning with class-specific augmentation improve CNN-based organic-inorganic waste classification

compared to a non-optimized baseline?

RELATED WORKS

Waste Classification and Deep Learning Approaches

Organic waste originates from biological material that decomposes naturally within a relatively short period (e.g., food scraps, garden waste), while inorganic waste consists of non-biological material such as plastic, glass, and metal that resists natural decomposition and requires recycling or thermal treatment [8], [9]. Accurate separation at the source is a prerequisite for effective integrated waste management [10].

Deep learning, a subset of machine learning based on multi-layer neural networks, learns hierarchical representations directly from raw data without manual feature engineering, making it well suited to unstructured data such as images [11], [12]. Convolutional Neural Networks (CNN) apply this principle to image data through local connectivity and parameter sharing: convolutional layers detect local patterns such as edges and textures, pooling layers reduce spatial dimensionality while preserving salient information, and fully connected layers perform the final classification [13], [14]. Within CNN architectures, ResNet-50 addresses the degradation problem encountered in very deep networks where added depth causes vanishing gradients and stalled learning in early layers through residual learning, in which each block learns a residual mapping relative to its input via a shortcut (skip) connection [13]. ResNet-50 comprises an initial 7×7 convolution, four stages of bottleneck blocks (3, 4, 6, and 3 blocks respectively; 16 blocks in total), and a global average pooling layer prior to classification, with approximately 25.6 million trainable parameters. Its ImageNet-pretrained weights provide feature representations that transfer efficiently to new visual domains with limited fine-tuning [13].

Transfer learning addresses data scarcity in a target task by reusing representations learned from a large source dataset such as ImageNet [15]. Two common strategies are feature extraction, in which the pretrained backbone is frozen and only a new classification head is trained, and fine-tuning, in which some or all pretrained layers are retrained with a small learning rate for deeper adaptation to the target domain; gradual unfreezing reopening blocks progressively from the top of the network is often used to stabilize this process [16]. Partial fine-tuning is generally most effective for small-to-medium datasets, as it preserves generic low-level features while adapting deeper, domain-specific ones [16]. Data augmentation complements this by artificially enriching the training distribution through geometric transformations (flipping, rotation, shearing, zoom) and photometric transformations (brightness, contrast, channel shift), reducing overfitting and improving robustness to real-world acquisition conditions [6], [17].

Comparison with Prior Studies

Table 1 summarizes the studies most directly relevant to this work's architecture, optimization strategy, and domain.

Table 1

Comparison of the Most Relevant Prior Studies

No	Author(s)	Title	Key Findings
1	Wu & Lin, (2022)	Effect of Transfer Learning on the Performance of VGGNet-16 and ResNet-50 for the Classification of Organic and Residual Waste	Compared VGGNet-16 and ResNet-50 with transfer learning on organic-versus- residual waste; ResNet-50 achieved the best performance after fine-tuning and parameter optimization.
2	Lin et al., (2022)	Applying a Deep Residual Network Coupling With Transfer Learning for Recyclable Waste Sorting	Developed RWNet-152, a ResNet-based model with transfer learning and augmentation, achieving high precision, recall, and F1-score for recyclable-waste sorting.
3	Yoo & Kang, (2023)	Class-Adaptive Data Augmentation for Image Classification	Proposed a class-adaptive augmentation policy — tailoring transformations to each class rather than applying a single uniform scheme — that outperformed conventional augmentation and improved CNN generalization.
4	Malik et al., (2022)	Waste Classification for Sustainable Development Using Image Recognition With Deep Learning Neural Network Models	Applied transfer learning with EfficientNet-B0 for solid-waste classification, reaching 85% accuracy with good computational efficiency; augmentation strategy was not examined in depth.

5	Yong et al., (2023)	Application of MobileNetV2 to Waste Classification	Used MobileNetV2 for domestic waste classification, reaching 82.92% accuracy with efficiency advantages over conventional CNNs.
6	Chatterjee et al., (2022)	IncepX-Ensemble: Performance Enhancement Based on Data Augmentation and Hybrid Learning for Recycling Transparent PET Bottles	Proposed an InceptionV3–Xception ensemble with transfer learning and augmentation, reaching 99.76% accuracy on transparent-PET-bottle classification.
7	Eltehewy et al., (2023)	Efficient Classification of Imbalanced Natural Disasters Data Using Generative Adversarial Networks for Data Augmentation	Used GAN-based augmentation for imbalanced data, improving accuracy to 88.5%.
8	Fu et al., (2021)	A Novel Intelligent Garbage Classification System Based on Deep Learning and an Embedded Linux System	Built a MobileNetV3-based real-time garbage-classification system with an average inference time of 0.63 seconds per image.
9	Sathishkumar et al., (2023)	Forest Fire and Smoke Detection Using Deep Learning-Based Learning Without Forgetting	Applied an Xception-based model with Learning-without-Forgetting, reaching 98.72% accuracy while retaining performance on prior data.

10	M. H. Saleem et al., (2022)	A Weight Optimization- Based Transfer Learning Approach for Plant Disease Detection of New Zealand Vegetables	Applied transfer learning with weight optimization and augmentation for plant-disease detection, reaching a mean Average Precision of 91.33%.
----	-----------------------------	---	---

As shown in Table 1, prior ResNet-based waste-classification studies [18], [19] demonstrate that transfer learning and fine-tuning are effective for organic/residual and recyclable-waste classification, but both apply a single, uniform augmentation scheme across all classes. Conversely, work that does adapt augmentation per class [20] was developed for general image classification rather than the waste domain. This study addresses that gap by jointly applying (1) gradual, partial fine-tuning of a ResNet-50 backbone and (2) a data augmentation scheme designed specifically for the differing visual degradation patterns of organic and inorganic waste a combination not previously reported for this dataset and architecture.

METHODOLOGY

Research Design

This study follows a sequential design comprising literature review, problem identification, dataset acquisition, preprocessing, class-specific augmentation, model implementation, staged fine-tuning, and evaluation, as illustrated in Figure 1.

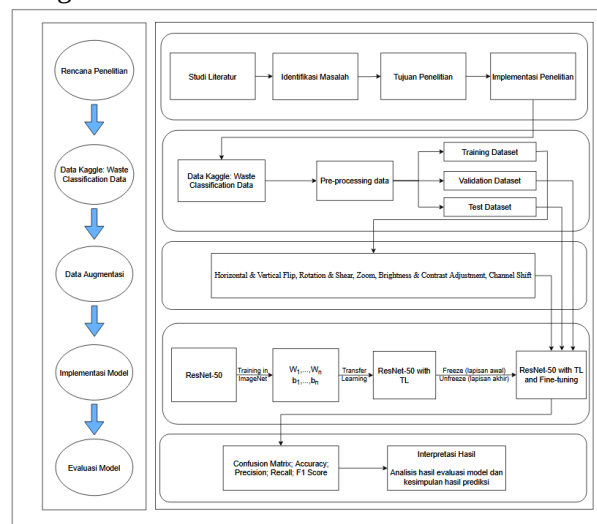


Figure 1. Research Framework

Dataset

The dataset used is Waste Classification Data, curated by Sashaank Sekar and publicly available on Kaggle (<https://www.kaggle.com/datasets/techsash/waste-classification-data>). It contains 25,077 labeled images across two classes: Organic (O), representing biological waste such as food scraps and

leaves, and Recyclable/Inorganic (R), representing materials such as plastic, glass, and metal. Although the dataset is provided pre-split into a training folder (22,564 images) and a testing folder (2,513 images), this study merged both folders and re-split the combined 25,077 images using stratified random sampling at a 70:15:15 ratio to preserve proportional class representation in every subset (Table 2).

Table 2

Dataset Split

Subset	Percentage	Number of Images
Training	70%	17,553
Validation	15%	3,762
Test	15%	3,762
Total	100%	25,077



Figure 2. Sample Images of the Organic and Inorganic Classes

The training set exhibited a mild class imbalance (~56:44 Organic-to-Inorganic ratio); class weights were computed accordingly, though not ultimately used during final training.

Preprocessing

All images were resized to 224×224 pixels with 3 color channels (RGB) using bilinear interpolation, chosen for its balance of image quality and computational efficiency relative to nearest-neighbor or bicubic methods. Pixel values were then normalized using ResNet-50's built-in `preprocess_input` function, aligning the input distribution with the scale used during ImageNet pretraining (Table 3).

Table 3

Preprocessing Parameters

Parameter	Value
Image size (IMG_SIZE)	224×224 pixels
Channels (IMG_CHANNELS)	3 (RGB)
Batch size (BATCH_SIZE)	32

Class Specific Data Augmentation

Augmentation was applied only to the training set and was designed differently for each class to match its visual degradation pattern:

Organic class: geometric transformations (horizontal/vertical flip, rotation, shear, zoom) combined with color/channel augmentation $\text{channel_shift_range} = 20.0$ to simulate decay-related color degradation, and a wider $\text{brightness_range} = [0.7, 1.3]$ to simulate extreme outdoor lighting conditions.

Inorganic class: geometric transformations only, with a narrower $\text{brightness_range} = [0.8, 1.2]$, since material color (e.g., of plastic, glass, metal) is a discriminative feature that should not be distorted.

This design forces the model to rely on texture rather than color alone for Organic samples, while preserving color as a reliable cue for Inorganic samples.

Model Architecture

The classification model uses a ResNet-50 backbone pretrained on ImageNet, with its original top layer replaced by a custom classification head. The convolution operation underlying ResNet-50's convolutional layers is defined as:

$$O(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot K(m,n) \quad (1)$$

where I is the input image, K is the learnable kernel, $O(i,j)$ is the output feature map at position (i,j) , i and j are spatial output indices, m and n are kernel offset indices, and $*$ denotes the convolution (cross-correlation) operator.

ResNet-50's core building block is the residual block, formulated as:

$$y = \mathcal{A}(x, \{W_i\}) + x \quad (2)$$

where y is the block's output, x is the block's input (identity shortcut), $\mathcal{A}(x, \{W_i\})$ is the residual function learned by the block's convolutional layers with weights W_i , and the term $+x$ is the shortcut connection added directly to the output. This design lets the network push $\mathcal{A}(x)$ toward zero when the optimal mapping is close to identity, easing optimization relative to learning the mapping from scratch.

The custom classification head consists of Global Average Pooling, a Dense layer (512 units, L2 regularization), Batch Normalization, Dropout (0.4), a second Dense layer (256 units, L2 regularization), Batch Normalization, Dropout (0.2), and a final Dense(1) output with sigmoid activation for binary classification.

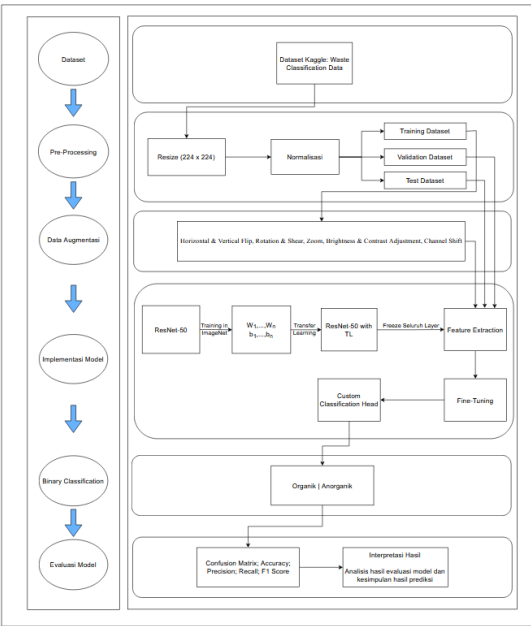


Figure 3. Model Architecture Diagram

Two Phase Training and Fine-Tuning Configuration

Table 4

Training and Fine-Tuning Configuration

Parameter	Value
Learning rate — Stage 1	1e-3
Learning rate — Stage 2	5e-5
Maximum epochs — Stage 1	15
Maximum epochs — Stage 2	40
Early stopping patience	10
ReduceLROnPlateau patience	4
Dropout rate	0.4
Unfreeze point	conv4_block1_1_conv
Optimizer	Adam
Loss function	Binary Crossentropy (label smoothing = 0.05)

Evaluation Metrics

Model performance was evaluated using a confusion matrix (True Positive/TP, True Negative/TN, False Positive/FP, False Negative/FN) and four derived metrics:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (3)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (4)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (5)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (6)$$

Ablation Design

To isolate the combined effect of the proposed optimization, two extreme scenarios were compared (Table 5): a Baseline model (fully frozen, randomly-initialized-behavior backbone with no fine-tuning and no augmentation) and the Proposed model (full two-phase fine-tuning and class-specific augmentation). Intermediate configurations (e.g., fine-tuning without augmentation, or augmentation without fine-tuning) were not tested; this is addressed as a limitation in Section 5.

Table 5

Ablation Scenario Definitions

Model	Fine-Tuning	Class-Specific Augmentation	Description
Baseline	Not applied	Not applied	Base model
Proposed	Applied	Applied	Fully optimized model

EVALUATION AND RESULTS

Baseline Model Evaluation

Table 6

Baseline Model Evaluation Results

Class	Precision	Recall	F1-Score	Support
Organic	0.8117	0.7881	0.7997	2,095
Inorganic	0.7431	0.7702	0.7564	1,667
Accuracy			0.7802	3,762

Macro avg	0.7774	0.7792	0.7781	3,762
Weighted avg	0.7813	0.7802	0.7805	3,762

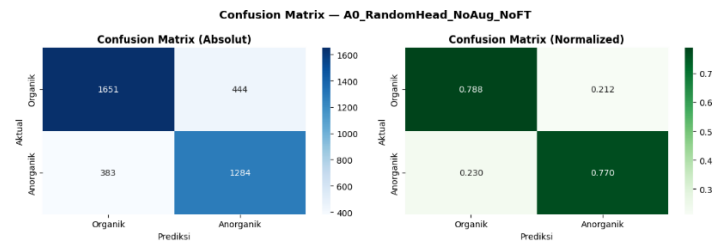


Figure 4. *Confusion Matrix Baseline Model*

Of 2,095 Organic test images, 1,651 were correctly classified and 444 were misclassified as Inorganic. Of 1,667 Inorganic images, 1,284 were correctly classified while 383 ($\approx 23.0\%$) were misclassified as Organic. Unlike the extreme single-class bias typically seen in unoptimized models, errors were fairly evenly distributed across both classes (Recall: Organic 78.81%, Inorganic 77.02%), indicating that without fine-tuning and augmentation the model had not learned a sharp decision boundary between the two classes.

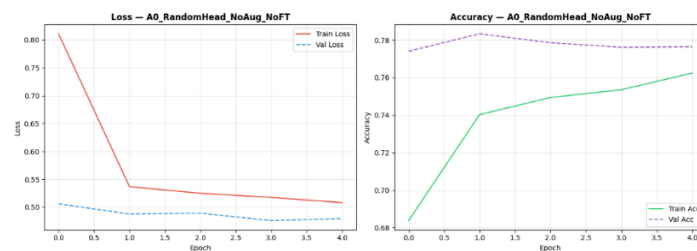


Figure 5. *Training/Validation Accuracy and Loss Curves Baseline Model*

Training accuracy rose gradually from about 0.685 to 0.762 by epoch five, while validation accuracy peaked early (≈ 0.784 at epoch one) before plateauing around 0.776–0.781. Validation loss similarly flattened early, from ≈ 0.505 to ≈ 0.483 . This early plateau confirms that a fully frozen backbone limits the model's learning capacity, regardless of additional training epochs.

Table 7

Proposed Model Evaluation Results

Class	Precision	Recall	F1-Score	Support
Organic	0.9799	0.9518	0.9656	2,095
Inorganic	0.9415	0.9754	0.9582	1,667

Accuracy		0.9623	3,762
Macro avg	0.9607	0.9636	0.9619
Weighted avg	0.9629	0.9623	0.9623

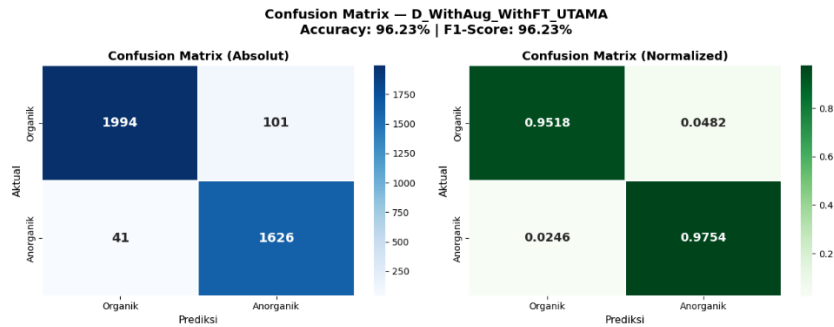


Figure 6. *Confusion Matrix Proposed Model*

Of 2,095 Organic images, 1,994 were correctly classified (101 errors). Of 1,667 Inorganic images, 1,626 were correctly classified (41 errors). Recall is now high and balanced across classes (Organic 95.18%, Inorganic 97.54%), showing the model is no longer biased toward either class as the baseline was.

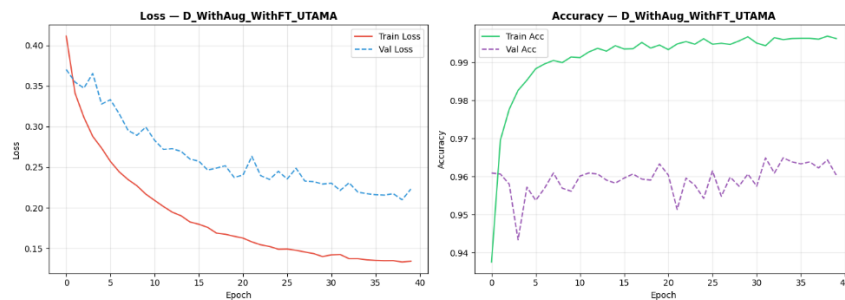


Figure 7. *Training/Validation Accuracy and Loss Curves Proposed Model*

Training accuracy climbed sharply in early epochs and approached 0.99 by the end of training, while validation accuracy stabilized around 0.95–0.97 with minor fluctuation. Both training and validation loss declined consistently with no strong sign of overfitting, though a small train validation gap is expected since class-specific augmentation makes the training and validation distributions not fully identical.

Qualitative Analysis of Misclassifications

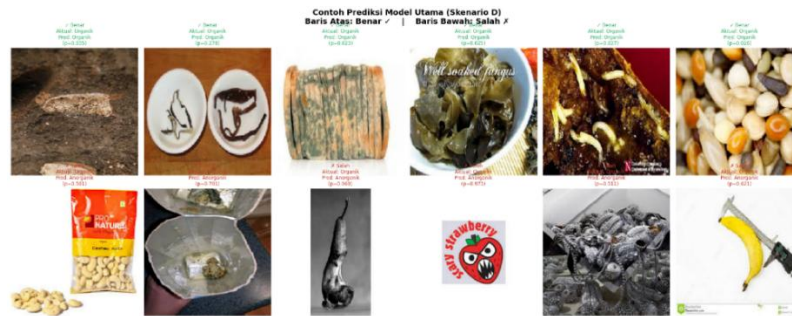


Figure 8. Misclassification Examples Organic Class

Visual inspection of Organic images misclassified as Inorganic revealed several recurring non-typical patterns: (1) organic items wrapped in plastic packaging, where packaging features visually dominate over the organic content (e.g., packaged nuts); (2) images rendered in non-standard color format (grayscale), removing color as a discriminative cue; (3) non-photographic, illustrated images rather than natural photographs, suggesting label noise in the source dataset; and (4) frames dominated by a non-organic object positioned alongside the organic item, such as a metal measuring tool. Notably, most of these misclassifications carried a high predicted probability ($p > 0.5$), indicating the model was confidently wrong rather than genuinely uncertain near the decision boundary.



Figure 9. Misclassification Examples Inorganic Class

Among the 41 misclassified Inorganic images, a somewhat different pattern emerged: (1) paper-based products such as tissues and tissue rolls, whose soft, fibrous texture visually resembles organic material despite being categorically inorganic due to industrial processing; (2) further non-photographic illustrations (e.g., a diagram of plastic-bottle cutting), consistent with the label-noise pattern seen in the Organic class; (3) frames where an organic element visually dominated despite the primary object being inorganic, such as a recycled-can flower vase surrounded by flowers and leaves; and (4) non-standard close-up images dominated by human hands or fingers (e.g., a beverage can being opened), reducing the visual prominence of the inorganic object itself. Most of these errors had a very low predicted Inorganic probability (below 0.3), again indicating confident rather than borderline mistakes.

Overall, both error groups suggest the model relies primarily on dominant surface texture and color cues within the frame rather than fully identifying the underlying material category a vulnerability on images with mixed visual composition, non-standard processing (grayscale, illustrations), or materially ambiguous products (paper-based items), which is an important consideration for dataset curation in future work.

Comparison with Baseline and a Reference Model

Table 8

Comparison of Baseline, and Proposed

Metric	Baseline	Proposed
Accuracy	78.02%	96.23%
Precision Organic	81.17%	97.99%
Precision Inorganic	74.31%	94.15%
Recall Organic	78.81%	95.18%
Recall Inorganic	77.02%	97.54%
F1-Score Organic	79.97%	96.56%
F1-Score Inorganic	75.64%	95.82%

The proposed model improved accuracy by 18.21 percentage points over the baseline (78.02% → 96.23%), with consistent gains in both classes: Inorganic recall rose by 20.52 points (77.02% → 97.54%) and Organic recall rose by 16.37 points (78.81% → 95.18%). Unlike the typical single-class bias seen in unoptimized models, the baseline here was already relatively balanced but uniformly low, so the fine-tuning and augmentation strategy is best interpreted as lifting overall discriminative capacity rather than merely correcting class imbalance.

Against the reference model of Wu and Lin (2022), which reported 96.60% accuracy using ResNet-50 with conventional transfer learning, the proposed model is broadly comparable ($\Delta = 0.37$ points) and even outperforms it on Organic recall (95.18% vs. 93.80%), Inorganic recall (97.54% vs. 91.60%), and Organic precision (97.99% vs. 97.30%), while the reference model performs better on Inorganic precision (94.50% vs. 94.15%) and on F1-score in both classes (97.80%/93.00% vs. 96.56%/95.82%). These differences are relatively small and may be influenced by dataset scale (22,010 images in Wu & Lin vs. 25,077 here), test-set composition, and differing augmentation strategies. This comparison should be read as indicative positioning relative to the literature rather than a formal statistical test, since the two studies use different datasets, data splits, and experimental conditions.

Two mechanisms plausibly explain the improvement over baseline: (1) gradual fine-tuning allows the deeper backbone layers to adapt to domain-specific visual features something unattainable when the entire backbone remains frozen, as in the baseline; and (2) class-specific augmentation enriches training variation to match each class's visual behavior color augmentation for Organic to simulate decay, and purely geometric augmentation for Inorganic to preserve material color as a discriminative feature. Together, these results empirically answer the research question: combining staged fine-tuning with class-specific augmentation significantly improves organic-inorganic waste classification performance over an unoptimized baseline, while remaining competitive with a conventional transfer-learning

approach reported in prior literature.

Model Parameter Complexity

Table 9

Estimated Model Parameter Count

Model Component	Estimated Parameters	Note
ResNet-50 backbone	23,587,712	ImageNet pretrained weights
Dense 512	1,049,088	$2048 \times 512 + 512$
Dense 256	131,328	$512 \times 256 + 256$
Dense 1 / sigmoid output	257	$256 \times 1 + 1$
Custom head total	$\approx 1,180,673$	Excludes Batch Normalization
Total model	$\approx 24,768,385$	Backbone + custom head

The custom classification head accounts for only about 4.77% of total model parameters, while the ResNet-50 backbone dominates over 95%. Consequently, in Stage 1 (feature extraction) only ≈ 1.18 million parameter under 5% of the full model are updated, while in Stage 2 (fine-tuning) the trainable parameter count increases substantially as the conv4 and conv5 blocks (94 layers, per the training log) are unfrozen. Given the depth of these residual stages, this explains why Stage 2 requires a much smaller learning rate ($5e-5$ vs. $1e-3$ in Stage 1) to avoid catastrophic forgetting of previously learned representations.

Convergence and Training Efficiency

Stage 1 reached its best weights at epoch 12 of a 15-epoch budget ($\text{val_loss} = 0.2825$) roughly 80% of the allotted epochs indicating relatively fast convergence. Stage 2, in contrast, reached its best weights at epoch 39 of a 40-epoch budget ($\text{val_loss} = 0.2098$), close to the upper bound of the epoch allowance, with validation loss declining steadily and consistently from 0.3701 at epoch 1 to 0.2098 at epoch 39, without signs of premature convergence or sudden overfitting. This indicates the Stage 2 learning rate ($5e-5$) was well calibrated small enough for stability, yet large enough to keep improving performance almost to the end of training. Early stopping (patience = 10) was never triggered through epoch 40.

This pattern suggests the proposed model may not have fully reached its optimal point within the epoch budget used in this study, meaning the reported 96.23% accuracy is potentially a lower bound rather than an upper bound on the architecture's capacity. Comparisons with the reference model (Table 8) and conclusions about optimization effectiveness should therefore be read as findings from a training run that had not fully converged, rather than as the architecture's maximum achievable performance.

Practical Implications and Deployment Considerations

The 96.23% accuracy achieved suggests the model is technically viable as an analytical component

within an automated, camera-based waste-sorting system for example, integrated with cameras at temporary disposal sites, waste banks, or a sorting conveyor belt functioning as a decision-support tool or first-pass sorting aid to reduce reliance on slow, error-prone manual sorting.

However, ResNet-50's ≈ 24.77 million parameters make it relatively heavy for edge deployment (e.g., Raspberry Pi, low-cost IoT cameras), likely requiring model compression such as pruning, quantization, or conversion to a lighter format (TensorFlow Lite, ONNX). Real-time inference latency and throughput on a conveyor belt or automated sorting arm were not measured in this study and warrant further testing. As a practical next step, latency can be measured by timing single-image predictions (e.g., using `time.time()/timeit` around `model.predict()` calls across tens to hundreds of test samples), and throughput can be measured as frames processed per second at a given batch size ideally benchmarked on both development hardware (GPU/CPU) and hardware closer to the intended deployment environment (e.g., CPU-only) to give a realistic picture of real-time feasibility.

CONCLUSION

This study successfully developed and optimized a ResNet-50-based CNN classifier for organic and inorganic waste images using transfer learning. Gradual, partial fine-tuning of the backbone's deeper layers, combined with a class-specific augmentation scheme color/channel augmentation to simulate organic decay and geometric-only augmentation to preserve inorganic material color improved model accuracy from a 78.02% baseline to 96.23%, an 18.21-percentage-point gain, with balanced improvement in both classes' recall (+16.37 points for Organic, +20.52 points for Inorganic). These results empirically answer the study's research question: combining staged fine-tuning with domain-adapted, class-specific augmentation significantly improves organic-inorganic waste classification performance relative to an unoptimized baseline, while remaining competitive with a conventional transfer-learning approach reported in prior literature. From an Information Systems perspective, the resulting model provides a reliable analytical component that can be integrated into a broader image-based waste-management information system, offering both a technical AI contribution and a foundation for automated waste-sorting information systems.

Several limitations should be considered when interpreting these results. First, the experimental design compares only two extreme ablation scenarios no optimization versus full optimization without intermediate configurations (e.g., fine-tuning alone, or augmentation alone); consequently, the individual contribution of fine-tuning versus class-specific augmentation to the 18.21-point accuracy gain cannot be empirically isolated from their combined effect. Second, although class weighting was computed to address the class imbalance identified in the training set ($\approx 56:44$ Organic-to-Inorganic ratio), it was not applied during final model training, so its specific contribution to the observed class-balanced performance cannot be separated from that of augmentation and fine-tuning. Third, the model was trained and evaluated exclusively on the Kaggle Waste Classification Data; its generalization to real-world field conditions such as those at waste banks or temporary disposal sites with variable lighting and background has not been empirically verified. Fourth, this study focused on optimizing accuracy and did not measure computational efficiency metrics such as inference latency, throughput, or model size, which are important considerations for deployment on resource-constrained edge devices.

Future research is encouraged to: (1) expand and diversify the dataset with field-collected images to improve real-world generalization; (2) extend the binary classification scheme into finer-grained multi-

class categories (e.g., plastic, metal, paper, glass) for more actionable recycling outcomes; (3) benchmark alternative transfer-learning backbones such as EfficientNet, MobileNetV2, or Vision Transformer for accuracy–efficiency trade-offs, particularly for edge deployment; (4) apply systematic hyperparameter optimization (e.g., Bayesian optimization, grid/random search) to the learning rate, unfreezing depth, and augmentation configuration; (5) integrate the model into a complete information system, such as a real-time, camera-based web or mobile application with a reporting dashboard to support operational decision-making at waste banks or disposal sites; and (6) evaluate the model on real hardware (e.g., IoT cameras linked to automated sorting mechanisms) to assess real-world inference speed and power consumption.

ACKNOWLEDGEMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

1. GENERATIVE AI DISCLOSURE

Generative AI tools were used to assist with reorganizing and translating content from the author's undergraduate thesis into this manuscript's structure, and with language editing for clarity and academic tone. All experimental design, data analysis, and conclusions are the author's own original work, and the author takes full responsibility for the accuracy, originality, and integrity of the submitted manuscript.

2. DATA AVAILABILITY STATEMENT

The dataset supporting this study, Waste Classification Data, is publicly available on Kaggle at <https://www.kaggle.com/datasets/techsash/waste-classification-data>.

3. AUTHOR CONTRIBUTIONS

Muhammad Luthfillah Andria conceptualized the study, curated and preprocessed the data, designed and implemented the model architecture and training strategy, performed the experiments and analysis, and wrote the manuscript. All authors read and approved the final manuscript.

4. CONFLICT OF INTEREST

The author declares no conflict of interest relevant to this work.

REFERENCES

- [1] D. S. Aulia, H. Arwoko, and E. Asmawati, “Klasifikasi Sampah Rumah Tangga Menggunakan Metode Convolutional Neural Network,” pp. 114–120, 2024, doi: 10.47002/metik.v8i2.956.
- [2] M. Malik, S. Sharma, M. Uddin, C. Chen, C. Wu, and P. Soni, “Waste Classification for Sustainable Development Using Image Recognition with Deep Learning Neural Network Models,” pp. 1–18, 2022.

- [3] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.
- [4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, and X. Zhai, "AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE," 2021.
- [5] W. Ting-Wei, H. Zhang, P. Wei, L. Fan, and he Pin-jing, "Applications of convolutional neural networks for intelligent waste identification and recycling," *Resour. Conserv. Recycl.*, vol. 190, 2023, doi: <https://doi.org/10.1016/j.resconrec.2022.106813>.
- [6] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
- [7] J. Jose, S. C. Mana, K. S. Babu, G. Kalaiarasi, and M. Selvi, "Enhancing waste classification accuracy with Channel and Spatial Attention-Based Multiblock Convolutional Network," *Environ. Monit. Assess.*, vol. 197, no. 2, p. 198, 2025, doi: 10.1007/s10661-025-13629-y.
- [8] F. Arifin, M. Habiburrakhman, and W. R. Gusti, "Classification of Organic and Inorganic Waste Types Based on Neural Networks," vol. 8, no. 1, pp. 78–85, 2023.
- [9] fahmi wafi Mubarakh, "Image Classification of Organic and Inorganic Waste Using Convolutional Neural Networks," pp. 781–783, 2024.
- [10] F. H. Puspitasari, S. Supriyadi, and M. Al-irsyad, "Analysis of Organic and Inorganic Waste Management Towards a Green Campus at Universitas Negeri Malang," vol. 44, no. Ismophs 2021, pp. 68–76, 2022.
- [11] S. Razavi, "Deep learning , explained : Fundamentals , explainability , and bridgeability to process-based modelling," *Environ. Model. Softw.*, vol. 144, no. August, p. 105159, 2021, doi: 10.1016/j.envsoft.2021.105159.
- [12] J. Egger, A. Pepe, C. Gsaxner, Y. Jin, J. Li, and R. Kern, "Deep learning — a fi rst meta-survey of selected reviews across scienti fi c disciplines , their commonalities , challenges and research impact," pp. 1–83, 2021, doi: 10.7717/peerj-cs.773.
- [13] M. A. Saleem, N. Senan, F. Wahid, M. Aamir, A. Samad, and M. Khan, "Comparative Analysis of Recent Architecture of Convolutional Neural Network," vol. 2022, 2022, doi: 10.1155/2022/7313612.
- [14] L. V. Haar, T. Elvira, and O. Ochoa, "An analysis of explainability methods for convolutional neural networks," *Eng. Appl. Artif. Intell.*, vol. 117, p. 105606, 2023, doi: <https://doi.org/10.1016/j.engappai.2022.105606>.
- [15] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning : Definitions , applications , advantages and limitations," *Expert Syst. Appl.*, vol. 242, no. June 2023, p. 122807, 2024, doi:

- 10.1016/j.eswa.2023.122807.
- [16] A. A. M. S. Ibrahim and J. R. Tapamo, "Transfer learning-based approach using new convolutional neural network classifier for steel surface defects classification," *Sci. African*, vol. 23, no. January, p. e02066, 2024, doi: 10.1016/j.sciaf.2024.e02066.
- [17] W. Sapitri, Y. N. Kunang, and I. Z. Yadi, "The Impact of Data Augmentation Techniques on the Recognition of Script Images in Deep Learning Models," vol. 8, no. 2, pp. 169–176, 2023, doi: 10.15575/join.v8i2.1073.
- [18] 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," vol. 10, no. October, pp. 1–11, 2022, doi: 10.3389/fenvs.2022.1043843.
- [19] K. Lin *et al.*, "Applying a deep residual network coupling with transfer learning for recyclable waste sorting," *Environ. Sci. Pollut. Res.*, pp. 91081–91095, 2022, doi: 10.1007/s11356-022-22167-w.
- [20] J. Yoo and S. Kang, "Class-Adaptive Data Augmentation for Image Classification," *IEEE Access*, vol. 11, no. March, pp. 26393–26402, 2023, doi: 10.1109/ACCESS.2023.3258179.
- [21] L. Yong, L. Ma, D. Sun, and L. Du Id, "Application of MobileNetV2 to waste classification," pp. 1–16, 2023, doi: 10.1371/journal.pone.0282336.
- [22] S. Chatterjee, D. Hazra, and Y. Byun, "IncepX-Ensemble : Performance Enhancement Based on Data Augmentation and Hybrid Learning for Recycling Transparent PET Bottles," *IEEE Access*, vol. 10, pp. 52280–52293, 2022, doi: 10.1109/ACCESS.2022.3174076.
- [23] R. Eltehwewy, A. Abouelfarag, and S. N. Saleh, "Efficient Classification of Imbalanced Natural Disasters Data Using Generative Adversarial Networks for Data Augmentation," 2023.
- [24] B. Fu, S. Li, J. Wei, Q. Li, Q. Wang, and J. Tu, "A Novel Intelligent Garbage Classification System Based on Deep Learning and an Embedded Linux System," vol. XX, pp. 1–13, 2021, doi: 10.1109/ACCESS.2021.3114496.
- [25] V. E. Sathishkumar, J. Cho, and M. Subramanian, "Forest fire and smoke detection using deep learning - based learning without forgetting," *Fire Ecol.*, 2023, doi: 10.1186/s42408-022-00165-0.
- [26] M. H. Saleem, J. Potgieter, and K. M. Arif, "A weight optimization-based transfer learning approach for plant disease detection of New Zealand vegetables," no. October, pp. 1–22, 2022, doi: 10.3389/fpls.2022.1008079.