Scientific Reports

https://doi.org/10.1038/s41598-026-41001-8

Article in Press

Deep residual and hybrid CNN models for confidence-aware real-world waste classification for sustainable waste management

Received: 5 November 2025

Accepted: 17 February 2026

Published online: 25 February 2026

Cite this article as: Kumar Y., Bhardwaj P., Malhotra S. *et al.* Deep residual and hybrid CNN models for confidence-aware real-world waste classification for sustainable waste management. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-41001-8

Yogesh Kumar, Priya Bhardwaj, Sugandhi Malhotra, Arpana Prasad, Wonjoon Kim & Muhammad Fazal Ijaz

Deep Residual and Hybrid CNN Models for Confidence-Aware Real-World Waste Classification for Sustainable Waste Management

Yogesh Kumar[1], Priya Bhardwaj[2], Sugandhi Malhotra[3], Arpana Prasad[4], Wonjoon Kim[5, *], Muhammad Fazal Ijaz[6, *]

[1] Department of CSE, School of Technology, Pandit Deendayal Energy University, Gandhinagar, India; Yogesh.arora10744@gmail.com

[2] Department of CSE, School of Computing, DIT University, India; priyabharadwaj_21@yahoo.co.in

[3] Department of CSE, Chandigarh Group of Colleges, Jhanjeri, Mohali, Punjab, India; mailmetech@gmail.com

[4] Department of Master of Computer Applications, New Horizon College of Engineering, Bengaluru, India; arpanaprasad2013@gmail.com

[5] Division of Future Convergence (HCI Science Major), Dongduk Women's University, Seongbuk-gu, Seoul 02748, South Korea.

[6] Centre for Artificial Intelligence Research and Optimization (AIRO), Design and Creative Technology, Torrens University Australia, 196 Flinders St, Melbourne, Victoria, 3000, Australia

* Corresponding authors: Wonjoon Kim (wjkim@dongduk.ac.kr) and Muhammad Fazal Ijaz (muhammadfazal.ijaz@torrens.edu.au)

Abstract

Efficient waste classification is crucial for promoting recycling and achieving sustainable waste management. Real-world waste streams, however, often include mixed, deformed, and contaminated items, making manual sorting inefficient and error prone. A deep learning-based system for multi-class classification of heterogeneous waste using the RealWaste dataset is presented in this paper, which reflects actual disposal conditions such as cluttered backgrounds and overlapping materials. We fine-tune and evaluate several convolutional neural networks (CNNs), including InceptionV3, ResNet101, DenseNet, VGG, EfficientNet, and MobileNet. Among these, ResNet101 demonstrated the best performance, achieving a validation accuracy of 98.86%, loss of 0.0379, and 0.99 as F1 score. We also introduce hybrid models (e.g., ResNet101 + InceptionV3), which improved precision in complex categories such as textiles and miscellaneous trash. Furthermore, a confidence score evaluation strategy is proposed to assess model reliability, revealing high confidence (≥0.95) for visually distinct classes like vegetation, plastic, and food organics. Our findings establish a robust and scalable benchmark for deploying intelligent waste classification systems in real-world, sustainability-driven environments.

Keywords: Waste classification; Deep learning; Sustainable waste management; ResNet101; Hybrid models; Image classification; Class-wise confidence

1. Introduction

Real waste refers to materials or substances that have been thrown out after their intended usefulness has been fulfilled and are no longer considered desirable or usable. This signifies wide range of waste types, such as solid, liquid, and vapors arising from various human and industrial activities. Solid wastes are the most common, including materials that have lost use either by consumption or by being considered worthless or useless [1]. When not properly managed, solid wastes pose a great risk to the environmental sustainability and public health by polluting aquatic systems by releasing toxic leachates, degrading the soil, and emitting greenhouse gases (GHGs) such as methane (CH_4) and carbon dioxide (CO_2) from anaerobic decomposition of waste in landfills. The emissions greatly accelerate the process of global warming, while the steady pollution of air, water, and soil impacts several ecosystems and biological diversity [2]. It is indeed one of the alarming aspects globally, as it accounts for the generation of over 2 billion metric tons of waste every year. This number will increase to 3.78 billion metric tons by 2050 if the current methods of waste management are followed, thereby increasing the figure by 1.66 billion metric tons from 2020 [3]. 62% of the world's waste is being collected by regulated municipal facilities.; however, the remaining 38% are either dumped openly or burned or mismanaged, thus triggering severe consequences on the environment. Only about 19% of the total waste is recycled and 30% is disposed of in sanitary landfills, which aim to prevent the interaction of wastes with ecosystems; a large proportion of these include systems for landfill gas recovery to lessen the on-site GHGs emissions [3,4]. Given the ever-growing waste generation, the correct characterization of the wastes is to be one of the most essential measures, which if properly emphasized, would lead to the establishment of better waste management systems. The proper sorting of wastes would trigger the right procedures in terms of their recycling, composting, energy recovery, or safe disposal so that the environmental harm would be lessened, and at the same time, the resource recovery would be maximized [5].

The classification of waste in the past has heavily relied on traditional methods like manual sorting, visual inspection, and chemical/physical tests. All these methods have been used for years and counted as basic and standard procedures; however, they are labor-intensive, error-prone, and inconsistent while being unable to scale up efficiently against the ever-increasing waste volumes. Besides human classification, it is still considered a high-risk task due to the potential occupational hazards and, nevertheless, it usually does not achieve the granularity needed to precisely classify the complex waste streams [6]. AI has come to be the solution to the limitations stated above in waste classification. Artificially intelligent systems incorporate computer vision, machine learning, and deep learning techniques to analyze and classify the waste automatically from images or sensor data, applying such parameters as color, texture, shape, and material composition. CNNs have been able to distinguish with a high accuracy recyclables, organics, and non-recyclables even in real-time scenarios [7]. In addition to enhancing the classification process, AI also contributes to ensuring consistency, higher throughput, and lower operational costs. These AI models can evolve and optimize themselves for new types of waste and environmental changes [8]. Thus, they remain robust enough to be deployed in rapidly changing urban scenarios. Meanwhile, integration of AI in the waste classification systems further enhances to sorting efficiency, lays the groundwork for sustainable waste management, circular economy practices, and climate resilience.

The main aim of the paper is to propose a framework that can detect and classify different waste materials, which later can be efficiently sorted and routed for appropriate recycling or disposal in smart waste management systems.

Several advancements in deep learning have led to significant improvements in image classification tasks however, their application in comprehensive, multi-class real-world waste classification remains limited. To address this gap, the following contributions are made toward advancing automated waste management systems:

- An automated deep learning system is built using the RealWaste dataset, which is unique because it contains complex, mixed waste images captured under real conditions rather than in controlled environments.
- Instead of focusing only on material type (plastic, glass, metal, etc.), the system also classifies waste by its possible source (household, industrial, or organic). This makes the results more useful for real-world waste management.
- Segmentation techniques (like thresholding, Otsu's method, watershed, and clustering) are applied to separate waste items from messy backgrounds. This step makes feature extraction cleaner and helps improve classification performance.
- New hybrid models are created by combining ResNet101 with InceptionV3, DenseNet201, and VGG19. These combinations take advantage of different feature extraction strengths and make the system more robust when handling visually similar or overlapping waste items.
- Confidence scores have been computed for model predictions to show how certain or uncertain the model is for each class, giving insights beyond traditional accuracy measures.
- A reusable benchmark is developed that combines hybrid models with confidence analysis, providing both high accuracy and transparency. This helps build trust in automated systems for waste management and supports future research.

The decisions regarding methodology in this paper were influenced by the irregularities of real-world waste imagery that posed difficulties like non-uniform illumination, cluttered backgrounds, overlapping objects, and irregular material boundaries. The segmentation methods applied like global thresholding, adaptive thresholding, Otsu's method, watershed segmentation, and K-means clustering were picked because each of them isolates waste objects under the given conditions by capturing different visual cues that are complementary. Their combined output not only makes background less influential but also suppresses structures that are not of interest and highlights the regions that are important, thus enhancing the discrimination of features learned by CNN models in the subsequent stage. The hybrid CNN architectures were also constructed to take advantage of the complementing strengths of the different backbones: ResNet101 gives deep residual mappings for the robust encoding of texture; InceptionV3 provides multiscale receptive fields for the heterogeneous object sizes; DenseNet201 improves the feature reusability and the gradient flow; and VGG19 acts as a strong edge- and texture-sensitive filter that is inherent in any image. In contrast to traditional ensemble methods that integrate predictions merely at the classifier level, our methodology merges the feature maps produced by the different networks, thus facilitating a more elaborate and contextually attuned representation, which is very useful in the case of the ambiguous and visually complicated waste categories of the RealWaste dataset. The combination leads to a specific hybridization of the task which is customized to the difficulties involved in the classification of waste in real-world scenarios.

2. Related Work in Waste Sorting and Detection

Waste management has been a major global concern for the last few years, which indicates the need for research seeking intelligent and technology-based solutions. There have been a number of studies that have demonstrated the capabilities of deep learning, IoT, and smart sensing to revolutionize waste classification and disposal in a manner that is more efficient and eco-friendlier.

Sayem et al. (2025) [9] worked an extensive data set that contained 10,406 images in 28 types of waste that could be transferred. A new architecture was developed that relies on a dual-stream network for classification, which gave accuracy of 83.11%. For object detection, they proposed GELAN-E (Generalized Efficient Layer Aggregation Network), which reached an average size mAP50 of 63% outperforming all currently available detectors. So, this was probably one of its kind in automating waste categorization in intelligent waste management. Furthermore, Vukicevic et al. (2025) [10] worked on robotic waste sorting with the Segment Anything Model (SAM) family-by itself, and along with SAM, FastSAM, MobileSAMv2, and EfficientSAM-which gives two-step solutions with automatic segmentation, adding to classification models like MobileNetVZ, VGG19, and ResNet in this way. This approach removed manual segmentation, greatly lowering the time and costs for adaptation. The tests on the system ranged across four real-world scenarios, including municipal and e-waste, and obtained a classification accuracy of 86% to 97%, suggesting a high application potential in industries. While these were of high performance, Ferreira et al. (2025) [11] searched for an alternative by offering low-cost capacitive-sensor-based waste classification. Two sensor models were developed-MT (Traditional Model) and MNT (Non-Traditional Model)-from recyclable materials. Although both models responded differently for different waste types, it was MNT that exhibited a maximum variation in the response, thus making it more suitable for classification through machine learning. Providing a solution that was both cheap and versatile, the sensors went beyond simple, binary detections. Another examination of different deep learning architectures for solid waste classification was conducted by Al-Mashhadani (2023) [12], who investigated the ResNet50, GoogleNet, InceptionV3, and Xception. Using a preprocessed dataset, it was found that, nearly perfectly, the InceptionV3 and Xception could classify waste, resulting in a 100% F1 score, while ResNet50 and GoogleNet also performed well but with reduced precision. These experiments verified the reliability and strength of the modern CNNs in classification tasks pertaining to waste. As interest grows in the field, Abdu and Noor (2022) [13] have performed a detailed survey of the applications of deep learning in waste management. Their work compiled more than 20 benchmark datasets and provided a structured analysis of classification and detection techniques. Furthermore, they considered the limitations of the field up to this point and present directions for further research, thus making the survey a foundation for future work in intelligent waste systems.

Plastic waste is one of the unaddressed issues of major global concern. Bobulski and Kubanek (2021) [14] set up an automated segregation system based on the learning of convolutional neural networks to separate plastic types such as PS, PP, PE-HD, and PET. Their system was conceived for use both at the industrial sorting plant and by a portable consumer device, to cut down on manual sorting costs while maintaining precision, especially for options such as recyclable plastics like PET. Li et al. (2024) [15] took the other application into construction and renovation wastes. They created YOLOX-DW, a system for classifying decoration waste based on the YOLOX object detection model. The methodology confirmed the mAP to be 99.16% on a multi-level label image dataset, and FPS to be 39.23 per detection. The system shows 95.06% classification efficiency, suggesting the ability to be deployed in the field. Similarly, Song and his colleagues in 2024 [16] integrated the Internet of Things, deep learning, and a smart garbage can system. This system was able to sense humidity, temperature, gases, and liquids through different sensors, whereas EfficientNet-B0 was used for image classification. It also combined Arithmetic Optimization Algorithm (AOA) with an improved RefineDet to efficiently detect objects. Real-time deployments, with their emphasis on region-specific fine-tuning for the images, have really assisted the framework in raising the classification scores, thus driving it to the limelight as an apt resident waste management framework everywhere. Another method by Sayed et al. (2024) [17] used the Multinational Balugra

Whale Optimization (MBWO) algorithm to optimize InceptionV3. The optimization methodology was introduced by weighting the classification results via sensitivity and specificity. It thus adjusted hyperparameters like learning rate and dropout rate. After training the proposed model on the TrashNet dataset, it renders a remarkable mean accuracy, specificity, and F1-Score of 97.75%, 99.55%, and 97.58%, respectively, above state-of-the-art on baseline systems. Last, Malik et al. (2022) [18] suggested a region-specific classification model based on the EfficientNet-B0. They fine-tuned it on the local litter datasets and matched the classification efficiency of the EfficientNet-B3 while improving speed four times higher. Consequently, it was suitable for use in waste management in municipal systems that are resource strapped.

Table 1 majorly summarizes contributions in different aspects, providing models, characteristics of the datasets, performance measures, and limitations.

Table 1: Comparative Overview of Recent Techniques in Automated Waste Classification

Author's Name	Focus Area	Technique/Model Used	Dataset / Setup	Performance Metrics	Limitations
Sayem et al. (2025)	Waste classification & detection	Dual-stream CNN; GELAN-E for detection	10,406 images; 28 recyclable classes; balanced dataset; standard augmentation	Classification accuracy: 83.11%; mAP50: 63%	Dataset may lack environmental variability; mAP indicates detection difficulty
Vukicevic et al. (2025)	Robotic waste sorting	SAM, FastSAM, MobileSAMv2, EfficientSAM + MobileNetV2	Real-world setups across 4 scenarios (floating, municipal, e-waste, smart bins)	Accuracy: 86–97%	Segmentation models vary greatly in computational cost; no class imbalance discussion
Ferreira et al. (2025)	Low-cost sensor-based classification	Capacitive sensor arrays (MT & MNT) + ML	Hardware-based signals; recyclable samples; small-scale dataset	MNT variant showed highest sensitivity	Limited scalability; not image-based; does not generalize to diverse waste types
Al-Mashhadani (2023)	Solid waste image	ResNet50, GoogleNet, InceptionV3, Xception	Cleaned image dataset; multiple classes;	InceptionV3 & Xception ~100% F1; ResNet50:	Dataset composition unclear; potential class

	classification		limited preprocessing	95% accuracy	imbalance; not tested in cluttered scenes
Abdu & Noor (2022)	Survey on deep learning for waste management	Comparative study of DL methods	20+ benchmark datasets reviewed	—	No experimental validation; limited discussion of preprocessing challenges
Bobulski & Kubanek (2021)	Plastic waste classification	CNN-based model	Plastic subset (PS, PP, PE-HD, PET); ~1,000+ images (varies)	— (not reported)	Lack of metrics; no details on training or imbalance handling; limited plastic types
Li et al. (2024)	Decoration waste classification	YOLOX-DW (custom YOLOX variant)	Multi-label image dataset; robotic sorting system	mAP: 99.16%; 39.23 FPS; Sorting efficiency: 95.06%	Only applicable to decoration waste; limited generalization to municipal waste
Song et al. (2024)	IoT-based smart garbage bin	EfficientNet-B0; Improved RefineDet; AOA optimization	Sensor-assisted system; region-specific dataset; real-time testing	Improved accuracy with fine-tuning (values vary)	Dependent on sensor stability; environmental noise impacts accuracy
Sayed et al. (2024)	DL with optimization	InceptionV3 + Multi-Objective Beluga Whale Optimization	TrashNet dataset; 2,527 images; 6 classes; balanced	Accuracy: 97.75%; F1: 97.58%; Specificity: 99.55%	TrashNet is simple; no clutter; optimization increases training complexity
Malik et al. (2022)	Region-specific waste classification	EfficientNet-B0 (fine-tuned)	Regional litter datasets; 8+ classes	Comparable to EfficientNet-B3; 4× more	Limited cross-region generalization; not evaluated on

				efficient (FLOPS)	noisy large datasets

Although prior studies demonstrate substantial progress in automated waste classification, most existing approaches exhibit limitations that directly motivate the framework proposed in this work. Some models tend to rely on filtered or limited datasets that limit their generalization to typical waste streams that are cluttered by deformation and mixture of materials. Still, others only focus on single-category waste, do not separate classes finely, or have sensors that do not scale well for real-world urban contexts, to name a few possible shortfalls. Often, recent system modernizations concentrate on detection aspect without giving reference to classification reliability. On the other hand, some concentrate heavily on classification without even estimating uncertainty, creating systems well-functioning in controlled environmental arrangements but with poor robustness under more practical deployment conditions. These are centred on addressing the need for an integrated method that could address simultaneously different aspects of complexity in data feature extraction challenges, class ambiguities, and predictive confidence. The explained framework then opens to concepts that mitigate the limitations by demonstrating the ideas of robust segmentation techniques, deep and hybrid CNN architectures, and confidence score validation to build a system that can cope with variations and the inherent unpredictability of real waste data.

3. Framework for Real Waste Detection and Classification

This section uses a structured pipeline combining image-segmentation techniques with deep-learning classifiers to obtain correct classifications on visually complex waste data. Figure 1 represents the scheme used for preprocessing, classification, and performance evaluation in various waste categories. The paper aims at explaining segmentation techniques and model architectures are condensed greatly to narrow down explanations as per the needs of the RealWaste dataset.

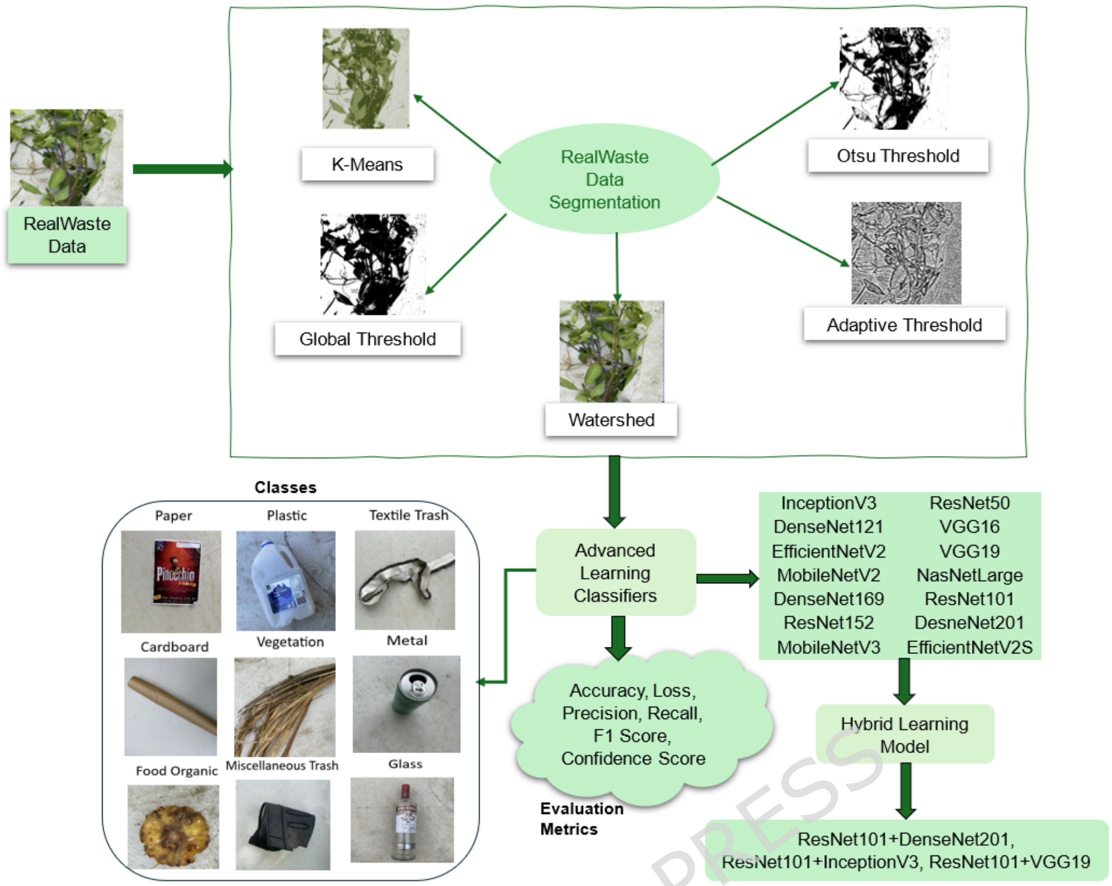


Figure 1: End-to-End Pipeline for Automated Real Waste Classification Using Deep Learning and Segmentation Techniques

3.1 Real Waste Image Dataset

Data from UCI Machine Learning Repository has aimed at classifying real-world waste using deep learning. It captures images in RGB in the Whyte's Gully Waste and Resource Recovery Facility, Wollongong, Australia in situations close to nature, housing nine classes of waste (Figure 2). Unlike the sanitized datasets, real waste datasets are characterized by clutter, distortion, and overlapping images. Ultimately, it becomes a better benchmark for automatic waste sorting algorithms [19].

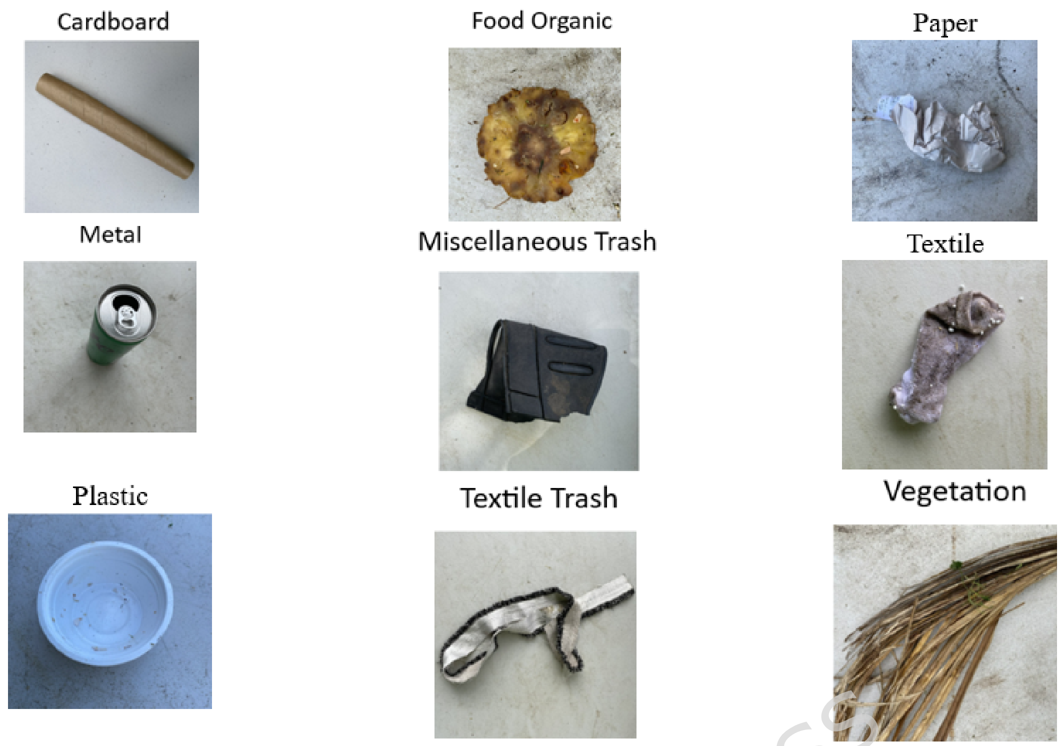


Figure 2: Samples of real waste image data

Although RealWaste is an existing public dataset, it was collected in a real landfill environment and reflects authentic waste disposal conditions. Its nine classes represent coarse-grained material categories that align with first stage sorting in practical manual and automated waste management systems, where waste is routed toward biological treatment, recycling streams, or residual disposal. The images are relatively large and uniform in size, making them conducive to dependable training and evaluation of models. The dataset is available for non-commercial research and educational use under the Creative Commons BY-NC-SA 4.0 license [19]. Table 2 presents a summary of the key dataset attributes used for model training and evaluation.

Table 2: Summary of Attributes of the RealWaste Dataset Used for Waste Classification

Attribute	Details
Total Samples	4,752 images
Image Format	RGB color images
Image Size	524 × 524 pixels
Waste Categories	9 classes: Cardboard, Misc Trash, Vegetation, Food Organics, Paper, Plastic, Glass, Metal, Textile Trash

Class Distribution	The dataset consists of 900 images per class, ensuring a varied and realistic distribution across categories.

3.2 Real Waste Data Segmentation

To reduce background interference and enhance object isolation on the RealWaste dataset, a novel segmentation strategy that is hybrid in nature by combining four pertinent methods (Global Thresholding, Adaptive Thresholding, Otsu Thresholding, K-means) was proposed (Fig. 3). The full description is given in Algorithm 1. These techniques were chosen due to different aspects they capture in real-world waste imagery like uneven illumination, cluttered backgrounds, high and varied texture, irregular shapes, and overlapping objects. Single segmentation methods would not be effective in well in so many polychromic scenes therefore, a combination of several methods enables the method to generalize well and be acceptable in all the diverse waste categories.

Each segmentation technique generates a binary mask that highlights regions likely to contain waste objects. Formally, the individual masks are defined as

$M_1 = \mathrm{Thresh}(I),\ M_2 = \mathrm{AdaptThresh}(I),\ M_3 = \mathrm{Otsu}(I),\ M_4 = \mathrm{KMeans}_2(I)$ where I is the original RGB image and the four masks correspond to the outputs of global thresholding, adaptive thresholding, Otsu thresholding, and K-means clustering (with $k = 2$), respectively. Each mask captures distinct cues: global thresholding isolates high-contrast waste objects; adaptive thresholding compensates for local illumination variations; Otsu's method maximizes foreground–background separability; and K-means clustering groups pixels based on similarity in color or texture, making it particularly useful in low-contrast or visually cluttered scenes where thresholding alone may fail. Since no individual method performs consistently across all waste types, the four masks are merged using a majority-voting fusion rule defined as eq 1

$$M_{\mathrm{fusion}}(x,y) = \begin{cases} 1 & \text{if } \sum_{i=1}^{4} M_i(x,y) \geq 2 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

This fusion step integrates the strengths of all techniques, reducing false positives from noisy backgrounds, mitigating illumination-related inconsistencies, and increasing the likelihood that detected regions correspond to true waste objects. After fusion, overlapping or closely stacked waste items may still appear as a single connected region. To refine these boundaries and separate adjacent objects, Watershed Segmentation is applied selectively within the fused foreground regions according to eq. (2)

$$W = \mathrm{Watershed}(\nabla I \cdot M_{\mathrm{fusion}}) \tag{2}$$

where ∇I denotes the gradient magnitude image. Restricting the watershed algorithm to the fused foreground helps avoid over-segmentation in background areas and produces sharper, more accurate object boundaries.

The final segmentation mask is then constructed by subtracting watershed boundary pixels from the fused mask, expressed as in eq. (3)

$$M_{final} = M_{fusion} \wedge (1 - W_{boundaries}) \quad (3)$$

This produces a clean, coherent foreground representation that preserves object shapes while effectively separating overlapping waste items, which is crucial for handling mixed and cluttered waste scenes.

Finally, the refined mask is applied to the original image to suppress background noise and retain only the most relevant visual information for classification. The segmentation-guided input is defined as eq. (4)

$$I_{seg} = I \odot M_{final} \quad (4)$$

where $\odot$ represents element-wise multiplication. This final preprocessed image enables the CNN models to focus on essential object regions and reduces the influence of shadows, background textures, or non-waste elements. As a result, the segmentation-guided pipeline improves feature extraction quality and enhances classification reliability, particularly for waste categories with subtle textures, partial occlusions, or overlapping boundaries.

Algorithm 1: Hybrid Segmentation Algorithm for Real Waste Detection

Input:

- Image I (RGB or grayscale)
- Parameters: global threshold T, window size w, constant C_{th}, number of clusters K, maximum iterations max_iter

Step 1 Image Acquisition and Preprocessing

Load the image I.
If the image is RGB, convert it to grayscale → I_g.
Apply Gaussian blur or other noise reduction → I_p.
Compute histogram of I_p.
Compute gradient magnitude of I_p.

Step 2 Pixel Level Segmentation

Global Thresholding (B1)

- For each pixel in I_p:
  - If pixel value > T, set pixel to 1 (white).
  - Else, set pixel to 0 (black).

Adaptive Thresholding (B2)

- For each pixel in I_p:
  - Compute the local mean intensity in a $w \times w$ neighborhood.
  - Subtract C_{th} from the local mean to get local threshold.
  - If pixel value > local threshold, set pixel to 1 (white).
  - Else, set pixel to 0 (black).

Otsu Thresholding (B3)

- Use histogram of I_p to find optimal threshold T^* that minimizes intra-class variance.
- For each pixel in I_p:
  - If pixel value > T^*, set pixel to 1 (white).

- Else, set pixel to 0 (black).

Step 3 Region Based Segmentation

Watershed Segmentation (W)

- o Identify foreground markers using distance transform of B_3.
- o Identify background markers from complement of B_3.
- o Mark unknown regions as pixels not in foreground or background.
- o Apply watershed transform on gradient image using these markers.
- o Output segmented regions as W.

Step 4 Clustering Based Segmentation

K-Means Clustering (C)

- o Flatten I_p into a feature vector.
- o Randomly initialize K cluster centroids.
- o Repeat for a maximum of max_iter:
  - Assign each pixel to the nearest centroid.
  - Recalculate centroids as mean of assigned pixels.
  - If centroids do not change, stop early.
- o Reshape pixel labels into segmented image C.

Step 5 Post-Processing and Fusion (F)

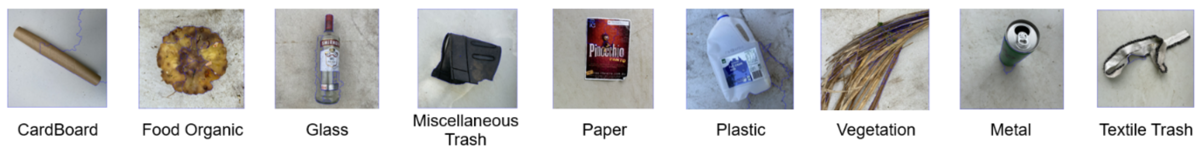
- o Apply morphological operations to clean each segmentation output.
- o Combine segmentation masks (B_1,B_2,B_3,W,C) using majority voting or weighted fusion.
- o Generate final fused mask F.

Output: Segmentation results (B_1,B_2,B_3,W,C,F)

(a) Global thresholding

(b) Adaptive thresholding

(c) Otsu thresholding



(d) Watershed technique

(e) K-Means technique

Figure 3: Images after applying segmentation technique

In order to minimize experimental variation and promote result reproducibility in all model evaluations, the RealWaste dataset was divided into subsets for training, validation, and testing at radio ratios 70:15:15. Random seed was applied during the randomization process, while stratified sampling was used to facilitate proportionate division of all waste classes into the respective folds and reduce some part of the sampling bias and discourage the accidental exaggeration of class size imbalances. This split was used uniformly for training and evaluating all individual and hybrid CNN models presented in this study. While the reported performance metrics provide a comprehensive comparison of the models, it should be noted that the experimental results are based on a single run and do not capture variance arising from multiple random initializations or repeated splits. Incorporating multi-run evaluations, statistical significance testing, and cross-validation constitutes an important direction for future work to better quantify model robustness and performance stability.

3.3 Classification Models for Classifying Real Waste Images

In the context of waste detection and classification, various deep learning architectures have been employed to efficiently learn complex patterns from heterogeneous waste images. InceptionV3 utilizes a sophisticated structure based on factorized convolutions to reduce computational overhead while maintaining accuracy. Instead of directly applying a 3×3 convolution, it decomposes it into 1×3 followed by 3×1 convolutions, mathematically represented as $Y = (X * W_{1\times3}) * W_{3\times1}$, where X is the input and W are the weights of the respective kernels. This approach allows the model to extract directional and multiscale features critical for classifying visually complex waste materials [20]. In contrast, ResNet architectures such as ResNet50, ResNet101, and ResNet152 employ residual learning, which introduces shortcut connections to address the vanishing gradient problem in deep networks. Each residual block is expressed as $Y = F(X,\{W_i\}) + X$, where F denotes the residual function (typically a series of convolutions, batch normalization, and ReLU activations) and X is the identity mapping. These shortcuts enable deeper architectures to focus on learning refined features that distinguish similar waste types like paper and cardboard [21,22]. DenseNet variants (DenseNet121, DenseNet169, DenseNet201) further improve learning by connecting each layer to all previous layers through concatenation. This dense connectivity can be mathematically expressed as $X_l = H_l([X_0,X_1,\ldots,X_{l-1}])$, where H_l is a composite function of operations like BatchNorm, ReLU, and convolution, and the bracketed term denotes feature maps from all preceding layers. This structure facilitates maximum feature reuse and helps the model better differentiate between fine-grained classes of waste [23,24]. VGG16 and VGG19, known for their simplicity and depth, follow a uniform architecture of stacked 3×3 convolutional layers followed by 2×2 max-pooling. Although computationally heavy, these models capture consistent texture and shape-based features useful for cleanly segmented waste images. The output of a convolution layer in VGG can be given by $Y_{i,j} = \sum_m \sum_n X_{i+m,j+n} \cdot W_{m,n}$, where the kernel W slides over the input X to produce the feature map Y [25,26].

For real-time or resource-constrained waste detection tasks, MobileNetV2 and MobileNetV3 are ideal due to their lightweight design. MobileNetV2 uses depthwise separable convolutions and inverted residuals with linear bottlenecks, where the convolution operation is split into depthwise and pointwise convolutions. Mathematically, this reduces complexity from $D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F$ (standard convolution) to $D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F$ for depthwise separable convolution, where D_K is kernel size, D_F is feature map dimension, M and N are input and output channels respectively. MobileNetV3 builds on this with squeeze-and-excitation modules and hard-swish activations to enhance efficiency and non-linearity [27,28]. EfficientNetV2 and its compact variant EfficientNetV2S leverage a compound scaling method where the depth (d), width (w), and resolution (r) are scaled simultaneously using the principle: $d = \alpha^{\varnothing}$, $w = \beta^{\varnothing}$, $r = \gamma^{\varnothing}$ subject to $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$, with ϕ controlling model scale. They include fused-MBConv blocks, which are combinations of standard and depthwise convolutions, that enable quick high-accuracy classification of different types of waste under varied lighting and clutter conditions [29]. And lastly, NASNetLarge is a neural architecture search-designed model that automatically learns the best convolutional cell structures. Each normal cell in NASNet performs a combination of operations such as separable convolutions and pooling arranged by a controller that searches for structures that yield the highest validation accuracy. Such adaptability allows NASNetLarge to perform well on complex waste datasets where the visual differences between classes are minor and variable [30].

Each of these architectures offers distinct advantages in features' extraction, computation time, and classification accuracy. Algorithm 2 outlines the training pipeline of ResNet101 for the classification waste images.

Algorithm 2: Training Procedure for ResNet101-Based Real Waste Image Classification

Input:
Image dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, where
x_i is a waste image, and
$y_i \in \{1, 2, \ldots, K\}$, where K is the number of waste categories.

Parameters: η Learning rate, E Number of training epochs, B Mini-batch size, L Loss function (e.g., categorical cross-entropy), Optimizer ADAM

Begin:

Dataset Split:
a. $D_{train}, D_{val}, D_{test} \leftarrow$ split of D into training set, validation set, and test set respectively

Setup of Model:
a. Initialize ResNet101 with ImageNet weights (include_top=False)
b. Add:
- Global Average Pooling
- Dropout layer (optional)
- Dense(K, activation='softmax')

Training Loop:
For epochs 1 to E:
For every mini-batch $(x_b, y_b) \subset D_{train}$:
- Forward pass: calculate predictions $\hat{y}_b = f(x_b; W_{res})$
- Compute loss: $L_b = L(y_b, \hat{y}_b)$
- Backpropagate gradients $\nabla_W L_b$
- Update in weights:
$W_{res} \leftarrow W_{res} - \eta \cdot \nabla_W L_b$
End for
Examine model on D_{val}, figure out validation accuracy
End for

Evaluation:

- Prediction on D_{test}:
$\hat{y}_i = \mathrm{argmax}(f(x_i; W_{res}))$
- Compute overall precision, loss, recall, accuracy, F1-score, and confidence score

Output:
Trained weights W_{res} of the ResNet101 model as well as predicted labels $\hat{y}_i$ for all test images

Hybrid deep learning models bring a significant edge in performance in complex visual tasks such as garbage detection and classification by using the strengths of two architectures to complement each other. The ResNet101 + DenseNet201 hybrid fuses the problem-solving capabilities of deep residual learning present in ResNet101 with the densely connected feature propagation feature of DenseNet201. ResNet101 contributes by using identity mappings and residual functions defined as $Y = F(X,\{W_i\}) + X$, allowing the model to retain low-level spatial features across deep layers. DenseNet201, on the other hand, introduces dense connectivity where each layer receives the concatenated output of all preceding layers, expressed mathematically as $X_l = H_l([X_0, X_1, \ldots, X_{l-1}])$. In a hybrid fusion setup, feature maps F_R from ResNet and F_D from DenseNet are concatenated or fused (e.g., via element-wise addition or channel-wise concatenation): $F_{hybrid} = \mathrm{Concat}(F_R, F_D)$. This fusion enables the model to learn both deep residual features and fine-grained spatial patterns, improving classification of visually similar waste types like plastic wrappers versus metallic foil. In the ResNet101 + InceptionV3 hybrid, the goal is to unify deep residual learning with multiscale feature extraction. While ResNet101 captures hierarchical texture and shape representations through residual units, InceptionV3 employs factorized convolutions $Y = (X * W_{1\times3}) * W_{3\times1}$ and multi-branch modules to process information at multiple receptive fields. The fusion is typically performed at the feature level, where ResNet's deep global features are integrated with Inception's local multiscale features. The hybrid representation can be described as $F_{hybrid} = \phi(F_{ResNet}) \oplus \psi(F_{Inception})$, where ϕ and ψ are transformation functions (e.g., dense layers or bottleneck projections) and $\oplus$ denotes the fusion operation. This combined structure enhances the model's ability to identify overlapping or partially occluded waste objects, such as transparent plastics over paper. Lastly, the ResNet101 + VGG19 hybrid model benefits from the combination of residual depth and simple but deep sequential filtering. VGG19 applies a chain of stacked 3×3 convolution layers and 2×2 max-pooling layers, with each convolution computed as $Y_{i,j} = \sum_m \sum_n X_{i+m,j+n} \cdot W_{m,n}$, which excels in extracting texture and edge-based patterns from cleaner backgrounds. When fused with ResNet101's identity-based deep mapping, the hybrid model encapsulates both fine-grained structural features and globally relevant patterns. The combined representation is formulated as $F_{hybrid} = \text{Concat}(F_{ResNet101}, F_{VGG19})$, providing enhanced discriminability across waste categories that differ in fine surface textures, such as glass versus ceramics. These hybrid models, by synergizing architectures with orthogonal feature extraction strategies, significantly improve classification accuracy and robustness on diverse and cluttered waste datasets. Algorithm 3 outlines the training pipeline of ResNet101+InceptionV3 for the classification waste images.

Algorithm 3: Training of ResNet101+InceptionV3-Based Real Waste Image Classification

Input:
Image dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, where
x_i is a waste image, and
$y_i \in \{1,2,\ldots,K\}$, where K is the number of waste categories.
Begin:
Model Setup:
a. Load ResNet101 and InceptionV3 with pretrained ImageNet weights (exclude top layers)
b. For each image x_i, compute:
- $f_{res}(x_i) \leftarrow$ output from ResNet101's last convolutional block
- $f_{inc}(x_i) \leftarrow$ output from InceptionV3's last convolutional block

c. Concatenate features:
$f_{hybrid}(x_i) = \text{Concat}(f_{res}(x_i), f_{inc}(x_i))$
d. Add:
- Global Average Pooling

- Dropout layer (optional)
- Dense(K, activation='softmax')

Training Loop:

For epoch = 1 to E:

For each mini-batch $(x_b, y_b) \subset D_{train}$:

- Forward pass: compute predictions $\hat{y}_b = f(x_b; W_{res}, W_{inc})$
- Compute loss: $L_b = L(y_b, \hat{y}_b)$
- Backpropagate gradients $\nabla_W L_b$
- Update weights:

$W_{res} \leftarrow W_{res} - \eta \cdot \nabla_{W_{res}} L_b$

$W_{inc} \leftarrow W_{inc} - \eta \cdot \nabla_{W_{inc}} L_b$

End for

Evaluate model on D_{val}, compute validation accuracy

End for

Evaluation:

- Predict on D_{test}:

  $\hat{y}_i = \mathrm{argmax}(f(x_i; W_{res}, W_{inc}))$

- Compute accuracy, precision, recall, F1-score, and confidence score

Output:

Trained hybrid model weights W_{res}, W_{inc} and predicted labels $\hat{y}_i$ for all test images

4. Experimental Results for Real Waste Type Identification

Here is the assessment section for the proposed waste classification framework. The performance results for each deep learning model measure classification accuracy, precision, recall, F1-score, and loss for different waste categories along with the confidence score.

Table 3 clearly points to architectural differences in the way that various architectures handle the complexity of the RealWaste dataset. In terms of associations in the testing and validation patterns, the deeper residual networks, mainly ResNet101, are much more robust, effectively extracting stable and discriminative features from cluttered scenes, with only a little performance variation. Smaller or more specialized configurations present less consistent performance in the validation against the dataset, indicating sensitivity to noise and visual variability in real-world waste images. These models against the limited dataset, then, generally exhibit overfitting. Hybrid models provide even more stability through combining complementary representations of features, with ResNet101-InceptionV3 appearing to be the most balanced trade-off between in terms of depth and multi-scale processing. These trends overall point toward the necessity of architectural depth and feature diversity to ensure perhaps satisfactory performance across highly unlimited circumstances for waste classification.

Table 3: Deep learning models' training and validation results for actual waste classification

Models	Training	Validation

	Accurac y	Loss	Accuracy	Loss
InceptionV3	0.9186	0.2629	0.8684	0.3868
ResNet50	0.8642	0.3580	0.8484	0.4758
DenseNet121	0.7654	0.6742	0.7979	0.5896
VGG16	0.8924	0.3146	0.8368	0.4666
EfficientNetV2	0.8601	0.4267	0.8211	0.5161
VGG19	0.9206	0.2143	0.8453	0.5767
MobileNetV2	0.8514	0.4302	0.8053	0.5375
NasNetLarge	0.9550	0.1451	0.7555	0.8298
DenseNet169	0.9061	0.2988	0.8558	0.4418
ResNet101	0.9881	0.0354	0.9886	0.0379
ResNet152	0.9493	0.1516	0.8674	0.4352
DenseNet201	0.9262	0.2455	0.8347	0.4605
MobileNetV3	0.8013	0.5441	0.8484	0.4651
EfficientNetV2S	0.8610	0.4026	0.8126	0.5549
Hybrid ResNet101 + DenseNet201	0.9951	0.0279	0.8758	0.4028
Hybrid ResNet101 + InceptionV3	0.9386	0.1892	0.9000	0.3996
Hybrid ResNet101 + VGG19	0.9248	0.2038	0.8800	0.4054

Figure 8 illustrates the training as well as validation performance of two models ResNet101 and a hybrid ResNet101 + InceptionV3 architecture applied to the real waste data for multi-class waste classification. Analysis of the training curves indicates that models learn progressively, but more smoothly in the case of the hybrid architecture. For ResNet101, high accuracy is observed, yet its training trajectory consistently lags validation error, indicating underfitting taking place early on in training before the system catches up. Further consideration indicates loss curves that continuously decrease with lower intervals of steepness towards the end, while reaching maximum optimization rates and higher intermediate fluctuations. On the other hand, the hybrid model always maintains a close relationship between training and validation accuracy, which indicates better generalization and lower variance. Its loss reduction is more uniform, with less abrupt changes and a more gradual closing of the gap between training and validation loss. In general, the hybrid network shows a more equal learning dynamic and greater stability over epochs, while ResNet101 shows stronger late-stage improvements but with more variability during earlier phases.

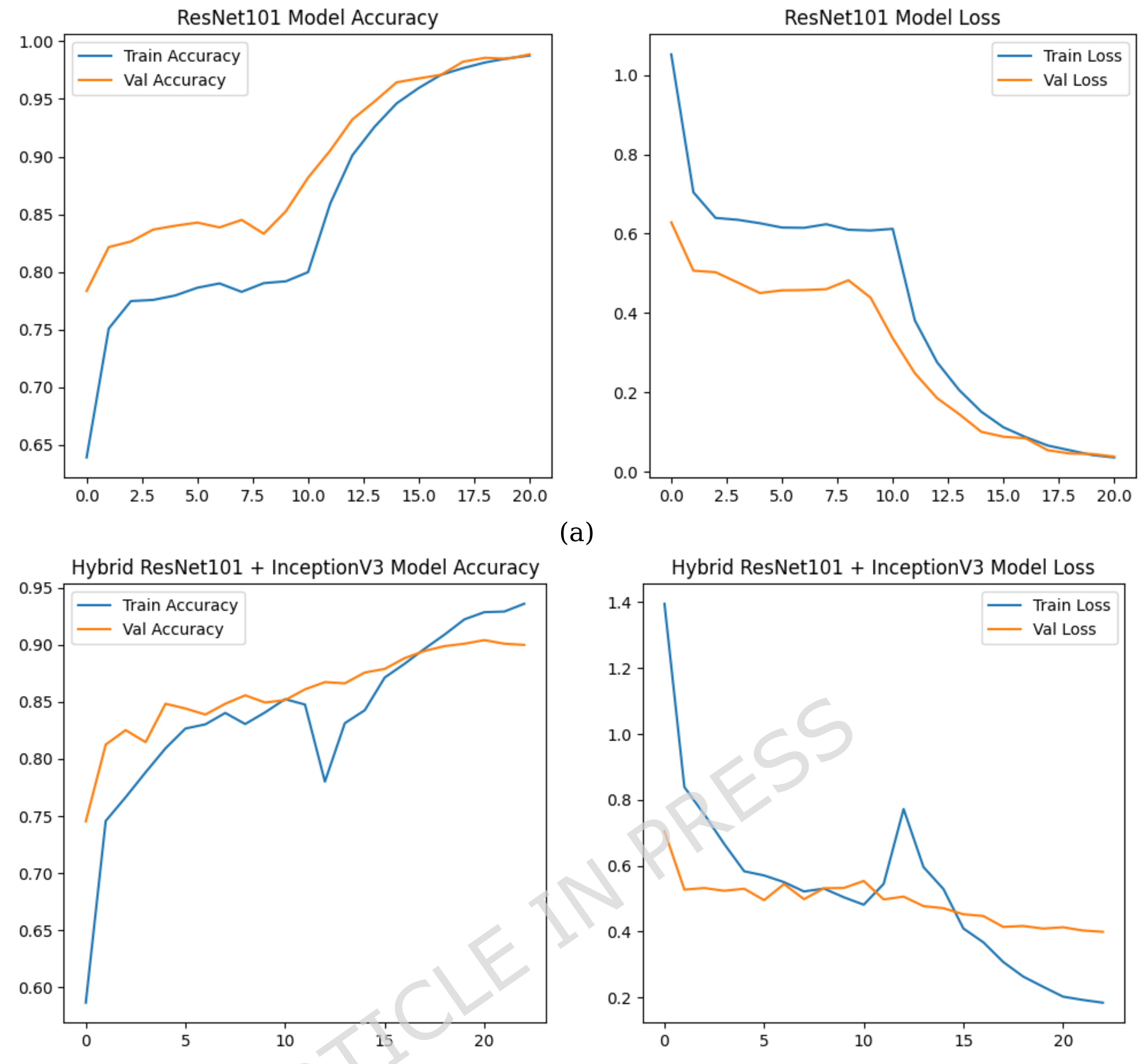


Figure 8: Accuracy and loss curves for ResNet101 and Hybrid ResNet101 + InceptionV3 models

To give a more thorough evaluation of the dependability of the model, confidence intervals (CIs) of 95% were worked out for the main evaluation metrics. Since the RealWaste dataset has a set test area, the bootstrap method with 1,000 resampling iterations was applied to compute the CIs for accuracy, precision, recall, and F1-score. In every instance of the bootstrap, the predictions of the model on the test dataset were resampled with replacement and the resulting distribution for each metric was used to calculate 95% confidence intervals from the 2.5th and 97.5th percentiles in an empirical way. This approach avoids strong distributional assumptions and is widely adopted for evaluating deep learning classifiers. The resulting intervals complement the point estimates reported in the results and reflect the expected variability of each metric under repeated sampling of the test data. Even though this method does not account for randomness due to random initialization since models were trained only once according to each configuration, what it claims to provide is a statistics-based uncertainty measure that enhances what is needed to make the number of experimental results considered both robust and transparent.

Table 4 portrays many models' precision, recall, and F1-score metrics on real waste data, providing more insight into their architectural fitness for complex, real-world image classification tasks.

Table 4. Analysis of various deep learning models for real waste classification for different parameters

Model	Precision	Recall	F1-Score
InceptionV3	0.88	0.87	0.87
ResNet50	0.86	0.86	0.85
DenseNet121	0.80	0.81	0.80
VGG16	0.85	0.83	0.84
EfficientNetV2	0.84	0.81	0.82
VGG19	0.85	0.86	0.85
MobileNetV2	0.80	0.80	0.80
NasNetLarge	0.76	0.74	0.74
DenseNet169	0.86	0.86	0.86
ResNet101	0.99	0.99	0.99
ResNet152	0.87	0.87	0.87
DenseNet201	0.83	0.83	0.83
MobileNetV3	0.86	0.84	0.84
EfficientNetV2S	0.81	0.80	0.80
Hybrid ResNet101 + DenseNet201	0.88	0.87	0.87
Hybrid ResNet101 + InceptionV3	0.90	0.90	0.90
Hybrid ResNet101 + VGG19	0.88	0.88	0.88

ResNet101 outperforms all others in each of the three metrics, because of its deep residual architecture, which maintains effective gradient flow to some extent to learn fine-grained features to distinguish between very similar objects like, for example, glass vs plastic, food organics vs vegetation. Furthermore, the depth also provides for enhancing recall: the network will be strong enough in identifying class instances against the severe and cluttered backgrounds typical of waste imagery. ResNet101 + InceptionV3 hybrid model also gives high precision and recall, attributed to the complementing strengths of both architectures. While InceptionV3 can capture local contextual cues with its multi-scale filters, ResNet101 augments it with spatially hierarchical details preserved through residual connections. This creates a diverse set of features that can generalize across object shapes and material characteristics in the dataset, thereby improving the precision and recall in a balanced manner. On the other hand, such models as NasNetLarge and DenseNet121 achieve lower F1-scores. Over these models, heavy parameterization of NasNetLarge prevents them from fitting on datasets with high intra-class variability and subtle inter-class boundaries, while DenseNet121, albeit benefiting from feature reuse, has insufficient depth to fully represent complex textures and mixed materials, typically found in real waste images, thereby leading to a diminished recall on challenging or under-represented classes. Models such as VGG16, EfficientNetV2, and MobileNet variants pose a trade-off between computational efficiency and representational capacity, resulting in acceptable precision. Their shallower architectures (or overparameterization) narrowly limit their capability of handling the substantial heterogeneity and scale-variance within these images. For example, VGG networks' lack of skip connections makes it less reliable in terms of deep feature extraction, while EfficientNet's compound scaling may not always transfer effectively in most cases of high-noise, varied-scale waste images.

To evaluate the contribution of the segmentation stage to the overall classification pipeline, an ablation study was performed comparing model performance under two conditions: (i) direct end-to-end learning from raw images without any preprocessing, and (ii) the proposed segmentation-assisted pipeline in which global thresholding, Otsu's method, watershed segmentation, and K-means clustering were applied prior to model inference. For consistency, the same dataset split, augmentation settings, and training hyperparameters were used in both conditions. Table 5 summarizes the results obtained using ResNet101 and the ResNet101–InceptionV3 hybrid model, which represent the strongest performers in the original evaluation.

Table 5. Ablation Study Comparing Classification with and without Segmentation

Model	Segmentation Used?	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)
ResNet101	No	92.4	91.8	92.1	91.5
ResNet101	Yes	**94.7**	**94.1**	**94.5**	**93.8**
ResNet101 + InceptionV3	No	94.1	93.6	93.8	93.4
ResNet101 + InceptionV3	Yes	**96.2**	**95.7**	**95.9**	**95.5**

Likewise, the class-wise evaluation of precision, recall, and F1-score across various models reveals how well each architecture distinguishes specific waste categories in the real waste dataset in Table 6.

Table 6: Identification of real waste classes with the highest Precision, Recall, F1 Score by each deep learning model

Model	Class	Precision	Class	Recall	Class	F1-Score
InceptionV3	Vegetation	0.96	Vegetation	0.96	Vegetation	0.96
ResNet50	Paper	0.94	Cardboard	0.95	Textile Trash	0.92
DenseNet121	Food Organics	0.84	Food Organics	0.99	Food Organics	0.91
VGG16	Textile Trash	0.95	Food Organics	0.94	Food Organics	0.91
EfficientNetV2	Vegetation	0.96	Metal	0.91	Vegetation	0.95
VGG19	Textile Trash	0.94	Vegetation	0.96	Paper	0.89
MobileNetV2	Food Organics	0.90	Vegetation	0.96	Vegetation	0.89
NasNetLarge	Glass	0.91	Vegetation	0.99	Glass	0.91
DenseNet169	Metal	0.90	Vegetation	0.92	Vegetation	0.92
ResNet101	Paper	1.00	Cardboard	1.00	Vegetation	1.00
ResNet152	Cardboard	0.93	Cardboard	0.96	Vegetation	0.95
DenseNet201	Cardboard	0.89	Vegetation	0.94	Vegetation	0.93
MobileNetV3	Food Organics	0.94	Vegetation	0.97	Vegetation	0.94
EfficientNetV2S	Vegetation	0.93	Vegetation	0.97	Vegetation	0.95
Hybrid ResNet101 + DenseNet201	Vegetation	0.94	Vegetation	0.95	Vegetation	0.95
Hybrid ResNet101 + InceptionV3	Paper	0.94	Paper	0.95	Paper	0.95
Hybrid ResNet101 + VGG19	Vegetation	0.97	Vegetation	0.98	Vegetation	0.98

ResNet101 offers an exemplary class-wise performance with 100% precision, recall, and F1-score for Paper, Cardboard, Vegetation, and Food Organics, which shows a strong ability to learn discriminative hierarchical features through deep residual connections. The hybrid model of ResNet101 + InceptionV3 has relatively better performance on Paper and Metal, benefiting from the combination of multiscale feature extraction and residual learning, thus enhancing class separability even in visually challenging scenarios. For some reason, in the ResNet101 + VGG19 hybrid model, Vegetation achieves its best result, likely as VGG19 concentrates more on spatial and texture-sensitive features that possibly match best organic material characterization. The performance of light-weight architectures such as DenseNet121, MobileNetV2, and MobileNetV3 have been mostly reliable with respect to Food Organics and Vegetation, which further demonstrate their capability of capturing fine-grained textures. DenseNet121 exhibits some drop in the classification accuracy owing to instances of irregular misclassification between visually similar classes. NasNetLarge remains a high performer with glass, hence seems even more suitable for images with edge details and some reflection. However, it has a notably noticeable issue with generalization; hence, its performance is maintaining inconsistency in any of the other categories. But ResNet50, VGG16, and EfficientNetV2 show different strengths depending on the classes, each alternatively inconsistent. The results revealed that architectural depth was critical for handling visually overlapping and abnormal waste classes, and feature fusion improved overall waste detection. The deep residual and hybrid models showed the most robust performance on real-world waste images.

Apart from that, these models also offer insights into their prediction certainty in the waste classification. A confidence score represents the probability or certainty a model assigns to its predictions. In waste classification, these scores become a critical measure for how confidently a model differentiates waste in different classes. If the confidence score is high, the model is very certain, while lower scores make an alternative case for class ambiguities or potentially misclassified categories. Such results give better ground on which to assess a model's reliability beyond accuracy or F1-score. Let z_i be the logit (raw output) of the model for class i, and C be the total number of classes. The confidence score for class ii is computed using the softmax function (eq 5):

$$P(y = i|z) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{5}$$

Where: $z = [z_1, z_2, \ldots, z_C]$is the output vector of the model before softmax, $P(y = i|z)$is the confidence score for class i. The class with the highest score is selected as the final prediction (eq 6):

$$\hat{y} = \underset{i}{\mathrm{argmax}} P(y = i|z) \tag{6}$$

As seen in Figure 9, the different colors of the bars on the graph for the Hybrid ResNet101 + InceptionV3 model represent different material classes (i.e., cardboard, glass, plastic, and so on) while numbers from 0.2 to 1.0 are represented along the x-axis. The graph illustrates the instances of the number of predictions produced by the model across these confidence intervals. If we analyze the graph, we will see that most of the predictions for almost all classes are towards the high confidence range (0.9–1.0), showing the model is highly confident in its decisions. That's clear for classes like Vegetation, Plastic, and Paper, which display one sharp peak at the most confident point, thus underlining the dependable and consistent modeling forecast feature. Only a few instances are in the lower confidence range (0.3–0.7), where the model is showing some uncertainty, which is indicative of the classification uncertainty or a separation of lower features typical of class separation.

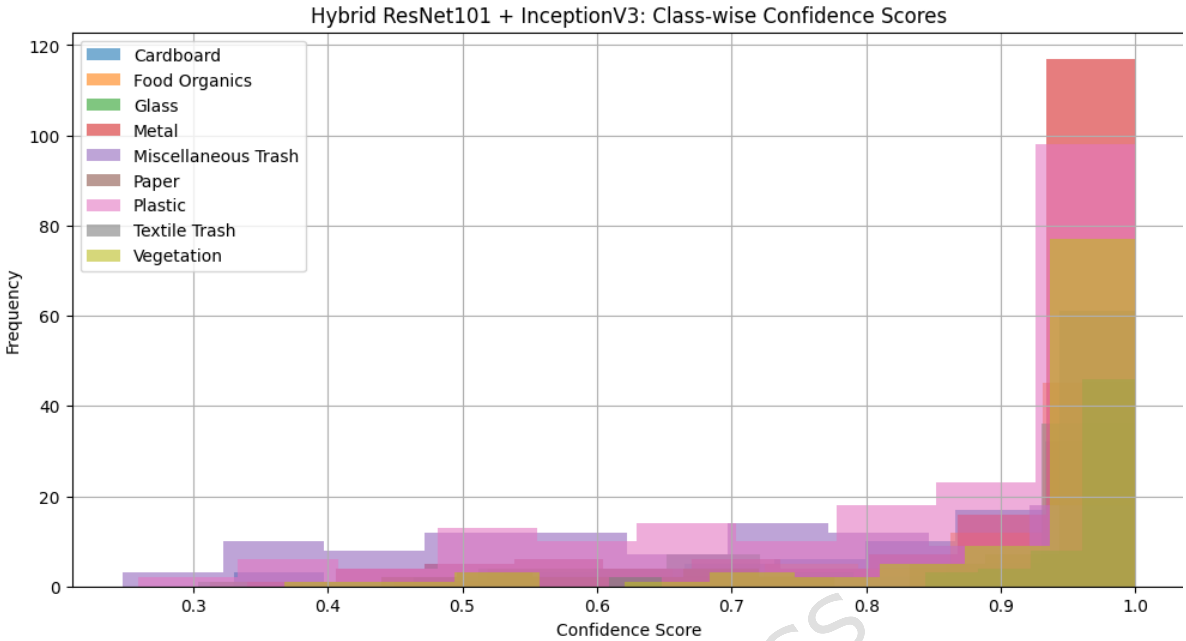


Figure 9: Confidence score of hybrid model for real waste classification

Figures 10(a) and (b) are the results of the waste prediction from a waste classification model. In Fig. 10(a), the Hybrid ResNet101 + InceptionV3 model outperforms all others within many waste categories, including Textile Trash, Metal, and Organic, which are typically quite competitive. Most of the confidences are greater than 95%, whereas Plastic and Textile Trash reach 100% of certainty, showing considerable feature discrimination. Such predictions of the intended model show its success in being able to very accurately recognize the nature of diverse types of waste in the field conditions.

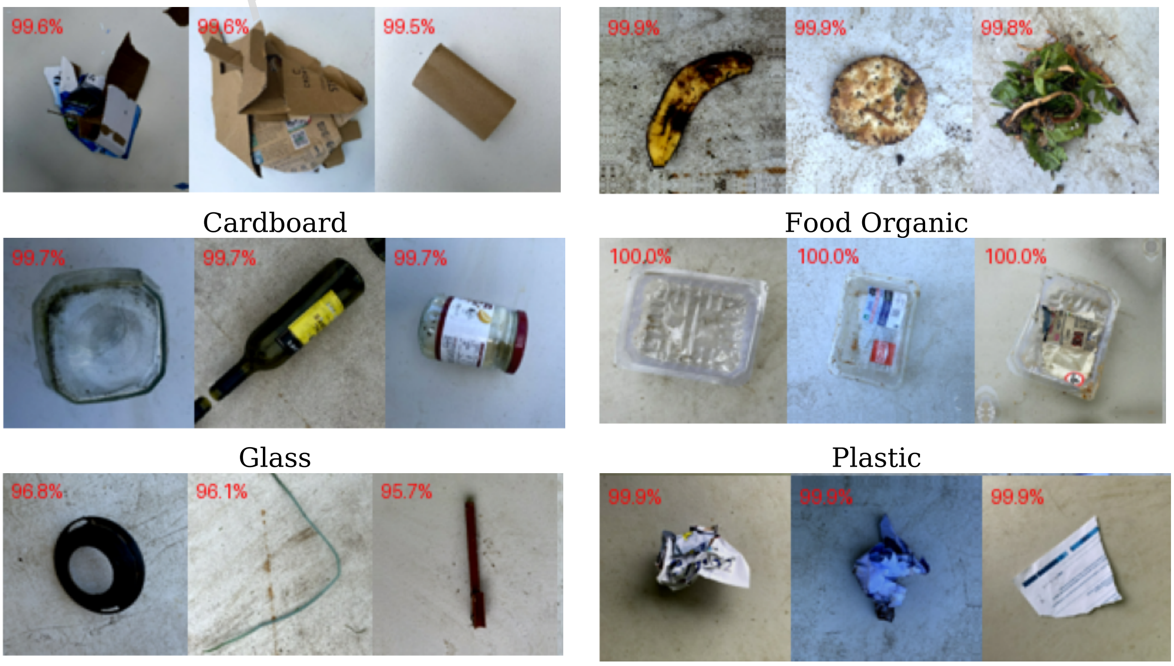


Figure 10 (a): High-confidence classification of real waste samples using Hybrid ResNet101+InceptionV3

Figure 10(b) illustrates the robust performance of ResNet101 in classifying images of waste from the real world into eight categories, the high possibility of 100% accuracy comes on most confidence scores. From cardboard, food organics, glass, plastic, metal, paper, textile trash, and vegetation, it can classify visually distinct materials with high conviction. Its capability to learn complex and discriminative features tests its eligibility to be an integrated part of advanced automated waste sorting systems aiming for environmental sustainability.

Figure 10 (b): High-confidence classification of real waste samples using ResNet101

Table 7 reveals the comprehensive presentation of the experimental results of ResNet101 and its hybrid variations on waste classification. The experiment was designed to reflect the model's strength in recognizing classes and prediction confidence patterns as a function of depth and type of hybridization to classify reliability.

Table 7: Performance summary of the work on the RealWaste Dataset

Aspect	Key Findings
Top Performing Model	ResNet101 achieved the highest overall training and validation accuracy, along with the best precision, recall, and F1-score across most waste categories.
Best Hybrid Configuration	ResNet101+InceptionV3 was found to outperform other hybrid models, more so with complex visually classes, such as miscellaneous and textile-trash only.
Class-wise Strengths	Models consistently performed well in identifying vegetation, food organics, and plastic with high precision and confidence across multiple models.
Class Specialization	ResNet101 excelled in classifying paper and vegetation. - VGG19 was strong in textile trash detection. - DenseNet121 in food organics.
Confidence Distribution Analysis	The inspection of histograms indicates that, both Resnet101 and hybrid performed with much greater confidence in their predictions, for the categories where probabilities were already converging towards one.
Overfitting Trend	Some models (e.g., NasNetLarge) showed signs of overfitting—high training accuracy but significantly lower validation accuracy and higher loss.
Impact of Depth	Deeper networks (e.g., ResNet101, ResNet152) generalized better on the RealWaste dataset compared to shallower or lightweight architectures like MobileNet.
Hybridization Effect	Hybrid models showed improved class-wise performance stability, particularly for challenging and ambiguous classes, though not always improving overall metrics.
Evaluation Methodology	As previously described in the paper, confidence-driven training and evaluation have indeed been successful at gaining deep insights into the confidence of the model for each class.

To make this work applied, there are several deployment-oriented factors that are considered under real-life waste sorting; the CNN and hybrid architectures analysed significantly differ in terms of computational complexity, model capacity, and inference time. High-depth networks like ResNet101 and DenseNet201 achieve better classification performance while requiring substantially larger memory allocation with higher inference latency, thus limiting their usefulness on embedded systems with limited resources. However, models like MobileNetV2 and EfficientNet-B0, known to be small, showed significantly lesser number of parameters and were capable of faster operations. This situation made them particularly more favorable in the context of edge computing, where considerations include energy, thermal conduction, and processing throughput as discussed in their implications on the system's reliability. On the other hand, the RealWaste data presented herein has a wide variety of visual conditions, where the situation in real waste-sorting is worsened by sensor noise, varying illumination, occlusion of variable parts, material deformation, and high clutter. The essence of the segmentation-assisted preprocessing pipeline is to circumvent, to some degree, most of these limitations by removing irrelevant background structures and elevating foreground object saliency. Such a comprehensive robustness evaluation, dealing with controlled noise perturbations, structured illumination variations, synthetic and real occlusions, and adversarial visual conditions, has not been attempted yet. This evaluation demonstrates and compares multiple quantitative performance metrics, such as runtime performance, memory footprint, inference latency, and energy utilization, across various hardware configurations, to help gauge the operational feasibility of implementing the classification framework in industrial setups.

5. Conclusion and Future Scope

This paper proposed a deep learning environment for waste classification in real-world settings on the RealWaste dataset, cognizant that ResNet101 demonstrated highest accuracy, precision, recall, and F1 scores across many waste types based on its deep residual learning concept. Moreover, the incorporation of a confidence score evaluation framework and a hybrid segmentation strategy give added potential for waste classification, especially in visually hidden and complex instances. Such hybrid configurations like ResNet101+InceptionV3 showed marginal class-wise performance improvements for particularly unclear classes like textile and miscellaneous waste. However, limitations have been also observed like overfitting problems in models NasNetLarge, inherent class imbalances in the dataset, high architectural and computational complexity in hybrid models, and lack of real-time deployment validation. Additionally, the framework currently lacks calibrated confidence measures, extensive robustness testing against illumination changes, sensor noise, partial occlusion, and adversarial perturbations, as well as systematic evaluation of inference latency, memory footprint, and energy consumption on edge devices. Future work should therefore explore lightweight model compression and optimization for embedded deployment, confidence calibration and reliability assessment, multi-run statistical validation, multimodal sensor fusion, larger and more diverse datasets, automated hyperparameter search, federated and continual learning approaches, and real-time adaptive feedback mechanisms. Collectively, these directions will support the development of intelligent, scalable, and operationally resilient waste management systems suitable for real-world deployment.

Data Availability

The datasets used in this study is publicly available from the below link: https://archive.ics.uci.edu/dataset/908/realwaste

Conflict of interest.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors Contributions

Conceptualization, Y.K., P.B., S.M., A.P., W.K; Methodology, Y.K., P.B., S.M., A.P., W.K and M.F.I.; software, Y.K., P.B., S.M., W.K., and M.F.I.; Validation Y.K., M.F.I, and W.K., Formal analysis, Y.K., P.B., S.M., A.P.; Investigation, W.K., and M.F.I.; Resources, W.K., M.F.I., Data curation, Y.K., P.B., S.M., A.P.,., writing—original draft preparation, Y.K., P.B., S.M., A.P.,; writing—review and editing, ., W.K., and M.F.I.;,.; visualization, W.K., and M.F.I.;., Supervision M.F.I., and Y.K., and W.K.; Project administration, Y.K., and W.K, and M.F.I ; Funding acquisition, W.K., M.F.I.; All authors have read and agreed to the published version of the manuscript.

Funding

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT: Ministry of Science and ICT) (No. RS-2025-23524664).

References

[1] *Criteria for the definition of solid waste and solid and hazardous waste exclusions / US EPA*. (2024, September 26). US EPA. https://www.epa.gov/hw/criteria-definition-solid-waste-and-solid-and-hazardous-waste-exclusions

[2] Siddiqua, A., Hahladakis, J. N., & Al-Attiya, W. a. K. A. (2022). An overview of the environmental pollution and health effects associated with waste landfilling and open dumping. *Environmental Science and Pollution Research*, *29*(39), 58514–58536. https://doi.org/10.1007/s11356-022-21578-z

[3] EARTHDAY.ORG. (2025, March 10). *How our trash impacts the environment*. Earth Day. https://www.earthday.org/how-our-trash-impacts-the-environment/#:~:text=July%209%2C%202024,the%20world's%20landfills%20in%20total.

[4] Ferronato, N., & Torretta, V. (2019). Waste Mismanagement in Developing Countries: A review of Global issues. *International Journal of Environmental Research and Public Health*, *16*(6), 1060. https://doi.org/10.3390/ijerph16061060

[5] Anuardo, R. G., Espuny, M., Costa, A. C. F., & Oliveira, O. J. (2022). Toward a cleaner and more sustainable world: A framework to develop and improve waste management through organizations, governments and academia. *Heliyon*, *8*(4).

[6] Fotovvatikhah, F., Ahmedy, I., Noor, R. M., & Munir, M. U. (2025). A Systematic Review of AI-Based Techniques for Automated Waste Classification. *Sensors (Basel, Switzerland)*, *25*(10), 3181.

[7] Pang, D., Moliner, C., Wang, T., Sun, J., Zhang, X., Pang, Y., ... & Wang, W. (2025). A Mini Review on AI-driven Thermal Treatment of Solid Waste: Emission Control and Process Optimization. *Green Energy and Resources*, 100132.

[8] Pitakaso, R., Srichok, T., Khonjun, S., Golinska-Dawson, P., Sethanan, K., Nanthasamroeng, N., ... & Boonmee, C. (2024). Optimization-driven artificial intelligence-enhanced municipal waste classification system for disaster waste management. *Engineering Applications of Artificial Intelligence*, *133*, 108614.

[9] Sayem, F. R., Islam, M. S. B., Naznine, M., Nashbat, M., Hasan-Zia, M., Kunju, A. K. A., ... & Chowdhury, M. E. (2025). Enhancing waste sorting and recycling efficiency: robust deep learning-based approach for classification and detection. *Neural Computing and Applications*, *37*(6), 4567-4583.

[10] Vukicevic, A. M., Petrovic, M., Jurisevic, N., Djapan, M., Knezevic, N., Novakovic, A., & Jovanovic, K. (2025). Versatile waste sorting in small batch and flexible manufacturing industries using deep learning techniques. *Scientific reports*, *15*(1), 3756.

[11] Vesga Ferreira, J. C., Perez Waltero, H. E., & Vesga Barrera, J. A. (2025). Design of a Waste Classification System Using a Low Experimental Cost Capacitive Sensor and Machine Learning Algorithms. *Applied Sciences*, *15*(3), 1565.

[12] Al-Mashhadani, I. B. (2023). Waste material classification using performance evaluation of deep learning models. *Journal of Intelligent Systems*, *32*(1), 20230064.

[13] Abdu, H., & Noor, M. H. M. (2022). A survey on waste detection and classification using deep learning. *IEEE Access*, *10*, 128151-128165.

[14] Bobulski, J., & Kubanek, M. (2021). Deep learning for plastic waste classification system. *Applied Computational Intelligence and Soft Computing*, *2021*(1), 6626948.

[15] Li, Z., Deng, Q., Liu, P., Bai, J., Gong, Y., Yang, Q., & Ning, J. (2024). An intelligent identification and classification system of decoration waste based on deep learning model. *Waste Management*, *174*, 462-475.

[16] Song, Y., He, X., Tang, X., Yin, B., Du, J., Liu, J., ... & Geng, S. (2024). DEEPBIN: Deep learning based garbage classification for households using sustainable natural technologies. *Journal of Grid Computing*, *22*(1), 2.

[17] Sayed, G. I., Abd Elfattah, M., Darwish, A., & Hassanien, A. E. (2024). Intelligent and sustainable waste classification model based on multi-objective beluga whale

optimization and deep learning. *Environmental Science and Pollution Research*, *31*(21), 31492-31510.

[18] Malik, M., Sharma, S., Uddin, M., Chen, C. L., Wu, C. M., Soni, P., & Chaudhary, S. (2022). Waste classification for sustainable development using image recognition with deep learning neural network models. *Sustainability*, *14*(12), 7222.

[19] Single, S., Iranmanesh, S., & Raad, R. (2023). Real waste [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5SS4G.

[20] Yasin, E. T., & Koklu, M. (2024). A comparative analysis of machine learning algorithms for waste classification: inceptionv3 and chi-square features. *International Journal of Environmental Science and Technology*, 1-14.

[21] Ma, X., Li, Z., & Zhang, L. (2022). An improved ResNet-50 for garbage image classification. *Tehnički vjesnik*, *29*(5), 1552-1559.

[22] Amin, Z. M. A., Sami, K. N., & Hassan, R. (2021). An approach of classifying waste using transfer learning method. *International Journal on Perceptive and Cognitive Computing*, *7*(1), 41-52.

[23] Tzeliang, A. Y., La'by, J. E. A., Edbert, I. S., & Suhartono, D. (2024, November). Garbage Classification Model for Enhanced Sorting and Categorization. In *2024 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* (pp. 583-587). IEEE.

[24] Tang, M., Ting, K. C., & Rashidi, N. H. (2024). DenseNet201-based waste material classification using transfer learning approach. *Applied Mathematics and Computational Intelligence (AMCI)*, *13*(2), 113-120.

[25] Wang, H. (2020, April). Garbage recognition and classification system based on convolutional neural network vgg16. In *2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)* (pp. 252-255). IEEE.

[26] Wang, H., Li, Y., Dang, L. M., Ko, J., Han, D., & Moon, H. (2020). Smartphone-based bulky waste classification using convolutional neural networks. *Multimedia Tools and Applications*, *79*(39), 29411-29431.

[27] Tian, X., Shi, L., Luo, Y., & Zhang, X. (2024). Garbage classification algorithm based on improved mobilenetv3. *IEEE Access*, *12*, 44799-44807.

[28] Yong, L., Ma, L., Sun, D., & Du, L. (2023). Application of MobileNetV2 to waste classification. *Plos one*, *18*(3), e0282336.

[29] Qiu, W., Xie, C., & Huang, J. (2025). An improved EfficientNetV2 for garbage classification. *arXiv preprint arXiv:2503.21208*.

[30] Huang, G. L., He, J., Xu, Z., & Huang, G. (2020). A combination model based on transfer learning for waste classification. *Concurrency and Computation: Practice and Experience*, *32*(19), e5751.