# AttenFusion-Net: An Attention-Enhanced Deep Learning Framework for Automated Municipal Solid Waste Segregation

Shaharior Islam Chowdhury * , Md Abdullah Al Azim , S. M. Fahim Abid , Saif Tasnim Chowdhury

Posted Date: 4 June 2026

doi: 10.20944/preprints202606.0372.v1

Keywords: municipal solid waste; waste classification; ResNet-CBAM; smart waste management; attention mechanism

 doi:10.20944/preprints202606.0372.v1



*Article*

# AttenFusion-Net: An Attention-Enhanced Deep Learning Framework for Automated Municipal Solid Waste Segregation

**Shaharior Islam Chowdhury [1,*], Md Abdullah Al Azim [2], S. M. Fahim Abid [3] and Saif Tasnim Chowdhury [4]**

[1] Bangladesh University of Business & Technology, Rupnagar, Mirpur-2, Dhaka, 1216, Bangladesh
[2] University of Global Village, Barisal, 8200, Bangladesh
[3] Islamic University of Technology, Dinbondhu Sen Road, Barisa, 8200, Bangladesh
[4] United International University, Madani Avenue, Badda, Dhaka, 1212, Bangladesh
* Correspondence: ishaharior@gmail.com

**Abstract**

The growing problem of municipal solid waste (MSW) generation presents a challenge in the efficient and scalable sorting of urban solid waste. Although deep learning models hold potential for automated waste segregation, their effectiveness is limited by the large Intra-class variance, Inter-class Similarity and Class Imbalance in some real world data. To overcome these limitations, we present a deep learning model with attention mechanisms, ResNet-CBAM, combining a ResNet-50 backbone with a Convolutional Block Attention Module (CBAM) to enhance feature representation through channel and spatial awareness. The approach is tested on the open Waste Segregation Image Dataset, which includes eight classes of biodegradable and non-biodegradable waste. To overcome the class imbalance problem, a combination of data augmentation and undersampling is applied. The findings reveal that our approach can attain an accuracy of **93.09%** and an F1 score of **0.9122**, surpassing some baseline approaches. The use of attention mechanisms improves the model's discriminative ability, resulting in better classification performance, especially in visually diverse environments. While minor errors still remain among visually similar classes, these results show a stable performance of the proposed model for all classes. In conclusion, the ResNet-CBAM model presents a reliable and transparent solution for waste classification, which can be further integrated into smart waste management and smart cities.

**Keywords:** municipal solid waste; waste classification; ResNet-CBAM; smart waste management; attention mechanism

## 1. Introduction

Urbanization has greatly altered the dynamics associated with the generation of waste throughout the world. With the rapid urbanization process, there has been an increasing amount of generation of municipal solid waste (MSW) due to which there have been increasing demands on the part of local authorities and organizations dealing with MSW. In the case of megacities, the rising population leads to generation of large amounts of MSW for which there is limited infrastructure to collect and process MSW, thus leading to complex situations regarding the issue of MSW management Shaibur et al. (2025). Poorly managed MSW has various repercussions including contamination of water and air, emission of greenhouse gases, and the permanent loss of resources that would otherwise be recycled Abogunrin-Olafisoye and Adeyi (2025). Source-level segregation is an essential component of any efficient MSWM system, whereby segregation is carried out at the time of generation into either biodegradable or non-biodegradable waste streams. While it is a fundamental process, manual segregation continues to be the common practice in most developing areas. This technique is not only expensive and laborious but also dangerous for those working there as it involves exposure to biological and chemical hazards Imam and Rafizul (2025).

 doi:10.20944/preprints202606.0372.v1



Advances in artificial intelligence (AI), machine learning (ML), and computer vision have opened promising pathways for transforming waste management. Image recognition and sensor-based systems can automatically detect, classify, and sort waste streams with minimal human intervention. Studies have demonstrated that AI-driven automation can meaningfully improve sorting efficiency, reduce operational costs, and make waste management more financially viable at scale Akhila et al. (2025); Imam and Rafizul (2025). The integration of such systems into Internet of Things (IoT)-enabled smart bins and collection networks represents a significant step toward intelligent, sustainable urban infrastructure. Within AI, deep learning, particularly convolutional neural networks (CNNs), has revolutionised image-based classification tasks. Deep learning models learn hierarchical visual features directly from data, eliminating the need for handcrafted feature engineering. Architectures such as VGG Ahmad et al. (2025), ResNet Potharaju et al. (2025), MobileNet Soni et al. (2024), and EfficientNet Tan and Le (2019) have each achieved strong performance across a broad range of visual recognition benchmarks. More recently, attention mechanisms and transformer-based architectures Wang et al. (2024) have raised the standard further, enabling models to selectively focus on the most informative regions of an image.

However, using deep learning for waste classification is not an easy task. There are two basic visual issues associated with the problem. The first is that of intra-class variance. This happens when objects within a certain type of waste have varying visual appearances despite being grouped under the same class. For example, a whole aluminum can and a squished aluminum can belong to the same class but are quite visually distinct from each other. The second issue is that of inter-class similarity, which means that different classes share similar visual features such as colors and textures Ahmad et al. (2025); Fang et al. (2023).

Compounding these challenges, real-world waste is rarely encountered in controlled conditions. Items may be occluded, partially degraded, contaminated, or photographed under variable lighting and backgrounds Thielmann et al. (2025). Most early AI-based systems were trained on clean, curated datasets that fail to capture this variability. As a result, models that perform well in laboratory settings often experience notable drops in accuracy when deployed in real-world environments. A further structural obstacle is dataset imbalance: natural waste streams are heavily skewed, with certain categories — such as food waste — vastly outnumbering others. Without explicit mitigation, this imbalance causes models to favour majority classes and perform poorly on minority ones, undermining both accuracy and practical utility. The absence of large-scale, well-annotated benchmark datasets for waste classification also impedes fair cross-study comparisons and hinders the field's collective progress.

Current approaches additionally tend to specialise in narrow waste taxonomies, limiting their generalisability across the diverse material streams found in urban settings. Many prior studies do not adequately address class imbalance, domain variability, or the computational constraints associated with real-time deployment. A further gap lies in the limited exploration of hybrid architectures that combine multiple attention mechanisms or fuse complementary feature representations. While recent fusion-based models have shown potential Ahmad et al. (2025), scalable and computationally efficient frameworks suited to resource-constrained environments in developing countries remain underexplored. To address these limitations, this paper proposes ResNet-CBAM, a novel attention-enhanced deep learning model for automated MSW classification. The architecture integrates a ResNet-50 backbone with a Convolutional Block Attention Module (CBAM), enabling the model to selectively emphasise informative channels and spatial regions while suppressing irrelevant background detail. To mitigate class imbalance, a combined strategy of data augmentation and random undersampling is applied to produce a balanced training set across all eight waste categories. Transfer learning and advanced augmentation techniques further address intra- and inter-class variance, improving the model's robustness to the visual diversity present in real-world waste streams.

The proposed framework is evaluated on the open-source Waste Segregation Image Dataset Dutt and Dutt (2022), which encompasses eight biodegradable and non-biodegradable waste categories.

 doi:10.20944/preprints202606.0372.v1



Experiments demonstrate that ResNet-CBAM achieves 93.09% accuracy and an F1-score of 0.9122, outperforming baseline architectures including ResNet-101, EfficientNet-B0, MobileNet-V3, and LSTM. Beyond standard metrics, the model is assessed using ROC-AUC, Cohen's Kappa, Matthews Correlation Coefficient (MCC), and Jaccard Index. Grad-CAM visualisations further confirm that the model attends to class-relevant image regions, supporting interpretability and transparency qualities essential for deployment in real-world systems.

The primary objectives of this study are as follows:

1. To develop and evaluate a comprehensive AI-driven system for automating municipal solid waste (MSW) segregation based solely on image data.
2. To employ state-of-the-art deep learning architectures to accurately classify waste materials into complex biodegradable and non-biodegradable categories.
3. To address dataset imbalance through advanced data augmentation and class-balancing strategies, ensuring robust learning across all waste categories.
4. To design a framework that is highly reliable, accurate, and capable of operating effectively in real-world urban environments.
5. To validate the practical value of the proposed ResNet-CBAM by benchmarking its performance against established baseline models and demonstrating its suitability for integration into sustainable smart-city infrastructures.

The main contributions of this work are summarized as follows:

- We develop ResNet-CBAM, a residual learning and attention-based approach for the classification of multi-class MSW.
- We address real-world challenges of class imbalance, intra-class variance, and inter-class similarity in waste images through a combined class-balancing strategy.
- We conduct in-depth experiments using both standard and advanced evaluation metrics, including accuracy, precision, recall, F1-score, ROC-AUC, Cohen's Kappa, MCC, and Jaccard Index.
- We perform an ablation study to quantify the independent and combined contributions of channel and spatial attention.
- We apply Grad-CAM visualisation to improve model interpretability and validate that learned features correspond meaningfully to waste-discriminative image regions.

The remainder of this paper is structured as follows. Section 2 reviews related work and motivates the research. Section 3 describes the dataset, preprocessing pipeline, and the ResNet-CBAM architecture. Section 4 presents and discusses experimental results. Section 5 addresses limitations and potential applications, and Section 6 outlines directions for future research.

## 2. Literature Review

The rapid development of artificial intelligence (AI) and deep learning technology has led to the improvement of intelligent waste classification systems, providing more accuracy and efficiency in municipal solid waste (MSW) sorting. The current body of work can be broadly classified into four categories: CNN-based waste classification approaches, object detection-based waste classification approaches, hybrid and ensemble approaches and IoT-based smart waste management systems. Although such frameworks have demonstrated promising performance, there are still some challenges with regard to generalization, robustness and efficiency.

### *2.1. CNN-Based Waste Classification Approaches*

Convolutional Neural Networks (CNNs) have been Convolutional Neural Networks (CNNs) have been widely used in waste classification tasks for their effective feature extraction ability. DenseNet, MobileNet, and VGG, which are based on transfer learning, have shown pramising results with high classification accuracy under laboratory conditions. For example, Yi and Kim (2024) achieved 95.2% accuracy with DenseNet121 but involved a limited number of samples, limiting its real-world

 doi:10.20944/preprints202606.0372.v1



applicability. Likewise, Sunardi and Fahmi (2023) reported a high accuracy by improving data preprocessing, but the model was sensitive to the quality of data and environmental conditions.

Other studies such as Akhila et al. (2025) and Gunaseelan et al. (2023) implemented VGG and hybrid residual networks with good performance. But these models are challenging to train due to lighting variations, occlusion and mixed waste streams. In summary, CNN-based solutions perform well in optimal conditions but are less robust in handling mixed waste streams in real-world scenarios.

### *2.2. Object Detection and Real-Time Approaches*

Various object detection methods (like YOLO and Mask R-CNN) have become very popular for real-time classification. These techniques enhance of object localization and allow integration in automated sorting systems. For instance, used a YOLO-based framework with moderate precision and mAP, whereas tested lightweight YOLO models for embedded systems, reaching accuracy of 93%.

Recent work has also explored attention modules in detection algorithms. Arishi (2025) combined Convolutional Block Attention Module (CBAM) with YOLOv8 to enhance detection; however, issues with small datasets and complex backgrounds persist. In addition, benchmarking studies, like Son and Ahn (2025), show the trade-off between speed and accuracy, with YOLO faster and Mask R-CNN more accurate. While effective, detection-based learning approaches can be computationally expensive and less effective under highly variable conditions.

### *2.3. Hybrid and Ensemble Learning Frameworks*

Hybrid and ensemble learning frameworks have been explored to enhance classification performance through the integration of various models or learning techniques. For example, Oise and Konyeha (2025) used a hybrid of EfficientNet and MobileNet to achieve high accuracy, while Pitakaso et al. (2024) used reinforcement learning combined with CNNs. While these methods improve classification accuracy, they typically also increase computational demands and diminish scalability.

Likewise, multi-stream and feature selection optimisation approaches have been proposed (Sayem et al., 2025; Wu et al., 2025) with increased accuracy. But these models often suffer from issues related to feature redundancy, computational efficiency and may not be suitable for real-time or low-resource scenarios.

### *2.4. IoT-Enabled Smart Waste Management Systems*

AI combined with Internet of Things (IoT) technologies has led to smart waste systems. Research like Alourani et al. (2025) has shown the potential for integrating deep learning with edge computing to enable real-time waste segregation. Similarly, there have been suggestions for using drones and robotics for waste segregation (Rajakumaran et al., 2023; Sirawattananon et al., 2021).

These approaches enhance automation and efficiency, but infrastructure and network dependency, along with deployment expenses, pose significant limitations. Additionally, they heavily depend on hardware and environmental factors.

### *2.5. Limitations of Existing Approaches*

Despite these advances, the current approaches for waste classification exhibit certain aspects of being limited:

- **Class imbalance:** Natural waste datasets are highly class-imbalanced, with majority classes (such as food waste) being overrepresented compared to minority classes, causing the model to be biased towards learning from the majority classes and perform poorly on minor ones (Castro-Bello et al., 2025; M. M. Islam et al., 2025).
- **Visual ambiguity:** Images of waste are visually noisy and ambiguous, with significant intra- and inter-class variability that makes distinguishing between similar-looking items like paper, plastic and food waste challenging (Ahmed Khan et al., 2024; M. M. Islam et al., 2025).

 doi:10.20944/preprints202606.0372.v1



- **Poor generalization:** Models trained on curated images often struggle to adapt to the complexities of the real world with background clutter and occlusions, and illumination changes (Jayaraman & Lakshminarayanan, 2024; Verber et al., 2026).
- **Computational efficiency:** State-of-the-art models, such as detection and hybrid models, have high computational costs, which affect their practicality for real-time or low-resource settings (Castro-Bello et al., 2025; Son & Ahn, 2025).
- **No attention:** CNN-based models often lack mechanisms to selectively attend to informative spatial and channel-wise features, limiting the model in distinguishing visually similar waste classes (Eryeşil et al., 2026; Rao et al., 2025).

### *2.6. Research Gap and Motivation*

From the above literature, it is clear that the current methods either focus on accuracy but are laboratory-based, or are real-time but lack accuracy. Very few methods consider class imbalance, visual ambiguity and variability of visual appearance in a holistic manner. Moreover, the use of attention mechanisms for improving feature representability is under-explored in multi-class classification of MSW.

To overcome these limitations, this research employs an attention-enhanced network, called **ResNet-CBAM**, integrating residual learning, channel attention and spatial attention mechanisms. To further enhance model performance, an imbalance-aware approach is adopted, which mitigates issues of under-represented classes. This holistic strategy aims to deliver a reliable and versatile waste segregation framework for real-world deployment.

**Table 1.** Summary of Representative AI-Based Waste Classification Studies.

| Year | Reference | Model / Approach | Performance | Key Limitations |
|---|---|---|---|---|
| 2025 | F. Islam (2025) | LSTM + NSGA-II + Blockchain-based framework | 27–35% material recovery, 41% emission reduction | High implementation cost, infrastructure dependency |
| 2024 | Yi and Kim (2024) | CNN (DenseNet121, MobileNetV2) | DenseNet121: 95.2% accuracy, F1: 0.95 | Small dataset (4,586 images), limited generalization |
| 2025 | Oise and Konyeha (2025) | Hybrid (EfficientNet + MobileNet + SNN) | 98% accuracy, AUC = 1.00 | Feature overlap, limited robustness to real-world variation |
| 2025 | Dipo et al. (2025) | YOLO-based detection model | 73% precision, 78% mAP | Moderate accuracy, poor performance in cluttered scenes |
| 2025 | Castro-Bello et al. (2025) | Lightweight YOLO variants | Up to 93% accuracy | High CPU usage, slow inference (1900 ms) |
| 2025 | Alourani et al. (2025) | VGG-19 + IoT-based system | 99.7% training precision | High cost, network dependency, scalability issues |
| 2023 | Rajakumaran et al. (2023) | Drone-based image classification | 95% accuracy | Environmental dependency, limited operational range |
| 2023 | Gunaseelan et al. (2023) | ResNeXt + ResNet-50 | 98.9% accuracy | High hardware cost, maintenance complexity |
| 2025 | Sayem et al. (2025) | Dual-stream deep learning model | 83.11% accuracy, mAP50: 63% | Moderate detection performance, clutter sensitivity |
| 2025 | Arishi (2025) | YOLOv8 + CBAM | 89.5% mAP | Dataset dependency, real-time constraints |
| 2025 | Son and Ahn (2025) | YOLOv8 vs Mask R-CNN | YOLOv8: 86.7%, Mask R-CNN: 91.2% | Speed–accuracy trade-off |
| 2025 | Akhila et al. (2025) | VGG16 + Web interface | 88% accuracy | Sensitive to lighting, occlusion, mixed waste |
| 2025 | Wu et al. (2025) | Optimization-based CNN (AHA variants) | 97.9% accuracy | Computationally intensive, scalability issues |
| 2023 | Mudemfu (2023) | YOLOv8 + EfficientNet-B0 | mAP: 96.5%, ROC: 98.4% | Heavy augmentation, real-world variability issues |
| 2024 | Chen et al. (2024) | YOLOv7-tiny + Edge computing | mAP@0.5: +1.7% improvement | Low FPS (9), limited real-time performance |

 doi:10.20944/preprints202606.0372.v1



## 3. Methodology

In this section, we provide the details of the experimental setting, including the dataset, data preprocessing, class balancing and the proposed ResNet-CBAM model. The workflow of the proposed system is depicted in Figure 1.

As the figure illustrates, there are four key steps in the pipeline: (i) data preprocessing and augmentation, (ii) dataset split and class-balance, (iii) feature extraction and classification using our proposed attention-enhanced deep learning model, and (iv) comparison to baseline models and evaluation measurements. This approach guarantees effective feature learning with methods that can deal with several real-life conditions such as imbalanced class distribution, and visual variability of waste types.

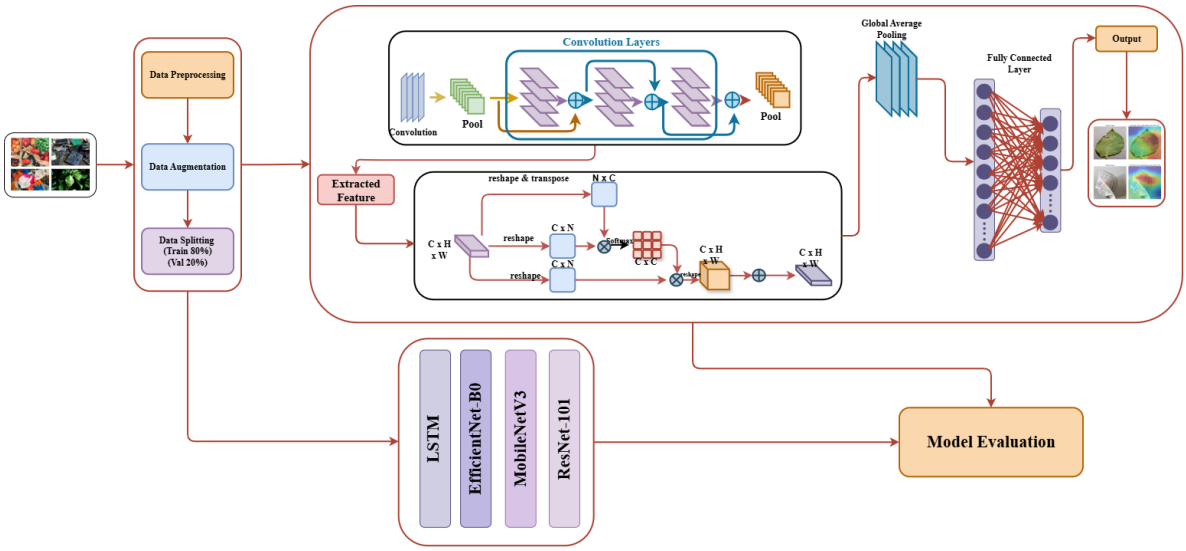


**Figure 1.** Overall workflow of the proposed ResNet-CBAM framework for waste classification.

### *3.1. Dataset Description*

This research uses the open-source Waste Segregation Image Dataset Dutt and Dutt (2022) for waste classification. There are labeled images from eight different waste classes, which are already split into a training (14,165 images) and validation (1,201 images) set to provide a standardized approach for model evaluation.

The classification task is a hierarchy of tasks. In the upper level, there are two broad categories: Biodegradable (Class 0) and Non-Biodegradable (Class 1). At the next level, each primary class is split into four sub-classes. The biodegradable class comprises food, leaf, paper and wood, whereas the non-biodegradable class is comprised of e-waste, metal, plastic bags and plastic bottles.

The class-wise distribution of the images in the training and validation sets is given in Tables 2 and 3, respectively.

**Table 2.** Summary of each class name and number of images from training set.

| Class Type | Class Name | Number of Images |
|---|---|---|
| Biodegradable (Class 0) | food_waste | 10066 |
| | leaf_waste | 1179 |
| | paper_waste | 860 |
| | wood_waste | 593 |
| Non-Biodegradable (Class 1) | e_waste | 180 |
| | metal_coins | 670 |
| | plastic_bags | 200 |
| | plastic_bottles | 417 |

 doi:10.20944/preprints202606.0372.v1



**Table 3.** Summary of each class name and number of images from validation set.

| Class Type | Class Name | Number of Images |
|---|---|---|
| Biodegradable (Class 0) | food_waste | 229 |
| | leaf_waste | 394 |
| | paper_waste | 212 |
| | wood_waste | 59 |
| Non-Biodegradable (Class 1) | e_waste | 55 |
| | metal_coins | 69 |
| | plastic_bags | 53 |
| | plastic_bottles | 130 |

### *3.2. Dataset Preprocessing*

The dataset was preprocessed to standardize and prepare the images for use in deep learning models. The images were uniformly resized to $224 \times 224$ pixels. The intensity of the pixels were scaled from $[0, 255]$ to $[0, 1]$ by a factor of $1/255$. This normalization improves numerical stability and accelerates convergence during training.

### *3.3. Data Augmentation*

To enhance the model's robustness and prevent overfitting, data augmentation techniques were used on the training data. Random rotations (up to 20°), width and height shifts (up to 10%), shear, zoom (up to 20%), and horizontal flips were applied in the augmentation process. These variations mimic the true distributions and diversity in the data.

The validation set was left unaugmented and only rescaled to ensure that model performance was not overestimated.

### *3.4. Class Imbalance Mitigation*

The dataset was oriented with marked class imbalance, where there were many more pictures of, for instance, food waste. To address this, class balancing was conducted (see Figure 2).

The goal was to achieve 800 images per class. For those classes with fewer than 800 images, extra samples were created by data augmentation until the class size was equal to 800. For the overrepresented classes, random undersampling took place. Thus, a balanced training set of 6,400 images (8 classes, 800 images per class) was obtained.

This approach helps balance the learning process across all classes, enhancing the performance of minority classes.

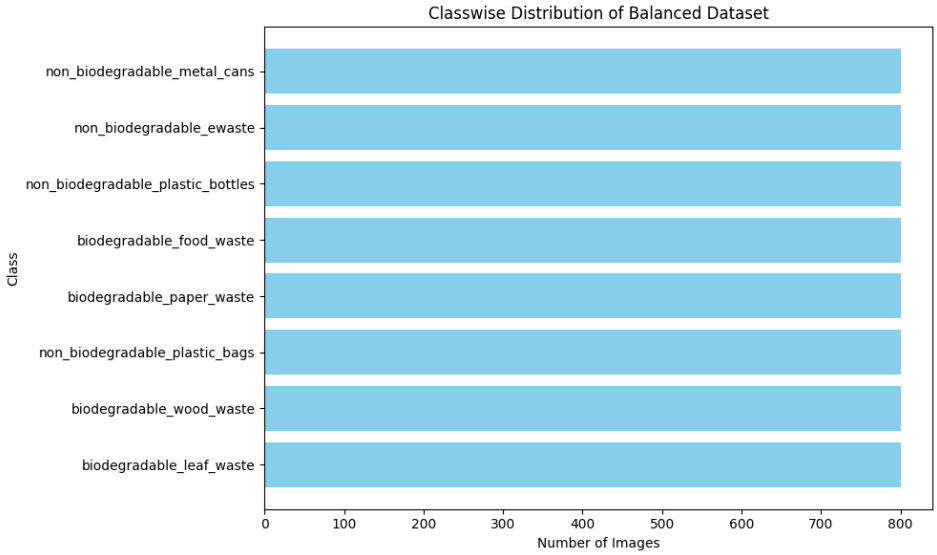


**Figure 2.** Proposed ResNet-CBAM Model Architecture.

 doi:10.20944/preprints202606.0372.v1



### 3.5. Dataset Partitioning

The original dataset was originally split into training (14,165 images) and validation (1,201 images) sets (roughly 92:8 ratio).

Following the class balancing, the size of the training set was modified to 6,400 images (800 images per class), with the validation set kept the same. This approach ensures that the model evaluation is done on an authentic imbalanced dataset.

Table 4 summarizes the dataset partitioning.

**Table 4.** Dataset Partitioning with training and validation sets.

| Dataset Partitioning | Number of images | Ratio of images |
|---|---|---|
| Training (before balancing) | 14,165 | 92% |
| Training (after balancing) | 6,400 | – |
| Validation | 1,201 | 8% |

### 3.6. Proposed Model: ResNet-CBAM Architecture

To overcome the issues of class imbalance, class similarity and changing real-world conditions in classification of municipal solid waste (MSW), this research presents an attention-based deep learning model called the **ResNet-CBAM**. This model combines a ResNet-50 CNN with Convolutional Block Attention Module (CBAM) to enhance feature extraction by incorporating adaptive channel and spatial attention.

Unlike traditional convolutional neural networks (CNNs) that equally weight all features, the proposed model adaptively highlights relevant features while down-playing less relevant background clutter. This is crucial in waste classification as items in this domain are high in intra-class variations and low in inter-class discriminations.

#### 3.6.1. Feature Extraction using ResNet-50

Given an input RGB image,

$$I \in \mathbb{R}^{H \times W \times 3}, \tag{1}$$

Deep feature maps are obtained with ResNet-50 backbone. The feature extraction process is:

$$F = \mathcal{R}(I), \tag{2}$$

where $\mathcal{R}(\cdot)$ denotes the ResNet-50 mapping, and

$$F \in \mathbb{R}^{C \times H' \times W'} \tag{3}$$

represents the resulting feature tensor.

In this paper, we use the CBAM module following the last residual block ($Conv5\_x$) of ResNet-50, for post-processing of higher-level semantic features before classification.

#### 3.6.2. Channel Attention Module

Channel attention is responsible for weighting features, to select what Global average pooling (GAP) and global max pooling (GMP) are used to obtain two descriptors:

$$F^c_{avg} = \mathrm{GAP}(F), \quad F^c_{max} = \mathrm{GMP}(F). \tag{4}$$

These are fed through a common multilayer perceptron (MLP) with a reduction factor of $r$:

$$M_c = \sigma\Big(W_2\,\delta(W_1 F^c_{avg}) + W_2\,\delta(W_1 F^c_{max})\Big), \tag{5}$$

where $\delta$ denotes the ReLU activation function and $\sigma$ is the sigmoid function.

 doi:10.20944/preprints202606.0372.v1



The refined feature map is obtained as:

$$F' = M_c \odot F. \tag{6}$$

3.6.3. Spatial Attention Module

The spatial attention module focuses on where important features are located. Channel-wise pooling is applied:

$$F^s_{avg} = \text{AvgPool}_c(F'), \quad F^s_{max} = \text{MaxPool}_c(F'). \tag{7}$$

The descriptors are concatenated and processed using a convolutional layer:

$$M_s = \sigma\Big(\text{Conv}_{7\times7}([F^s_{avg}; F^s_{max}])\Big). \tag{8}$$

The final refined feature map is:

$$F'' = M_s \odot F'. \tag{9}$$

3.6.4. Feature Aggregation and Classification

The refined features are aggregated using Global Average Pooling:

$$Z = \text{GAP}(F''), \tag{10}$$

followed by a fully connected layer and softmax activation:

$$\hat{y} = \text{Softmax}(WZ + b), \tag{11}$$

where $\hat{y} \in \mathbb{R}^K$ represents class probabilities.

3.6.5. Loss Function and Optimization

To address class imbalance, weighted categorical cross-entropy is used:

$$\mathcal{L} = -\sum_{i=1}^{K} \alpha_i y_i \log(\hat{y}_i), \tag{12}$$

where $\alpha_i$ represents class weights inversely proportional to class frequency.

The model is optimized using the Adam optimizer with a learning rate of $1 \times 10^{-4}$.

3.6.6. Model Complexity

The ResNet-CBAM model proposed in this work has a total of around 25.6 million parameters and 4.1 GFLOPs for a forward pass. CBAM adds a slight increase in computational cost, but enables more discriminative features to be learnt.

In summary, the combination of residual connections and attention modules allows our model to efficiently learn discriminative features, resulting in better classification accuracy in real-world scenarios.

 doi:10.20944/preprints202606.0372.v1



**Algorithm 1** Training Pipeline of the Proposed ResNet-CBAM Framework.

**Require:** Training dataset $\mathcal{D} = \{(I_i, y_i)\}_{i=1}^{N}$, learning rate $\eta$, batch size $B$, epochs $E$
**Ensure:** Trained ResNet-CBAM model

Initialize ResNet-50 backbone parameters $\theta$
Initialize CBAM parameters $\phi$
Initialize Adam optimizer with learning rate $\eta$
**for** epoch $= 1$ **to** $E$ **do**
  **for** each mini-batch $(I, y)$ of size $B$ **do**
    *// Data Preparation*
    Resize input images to $224 \times 224$
    Normalize pixel values to $[0, 1]$
    *// Forward Propagation*
    Extract feature maps: $F = \mathcal{R}(I)$
    Compute channel attention map:

$$M_c = \sigma\big(MLP(GAP(F)) + MLP(GMP(F))\big)$$

    Apply channel attention: $F' = M_c \odot F$
    Compute spatial attention map:

$$M_s = \sigma\big(\text{Conv}_{7\times7}([AvgPool(F');\ MaxPool(F')])\big)$$

    Apply spatial attention: $F'' = M_s \odot F'$
    Aggregate features via global average pooling: $Z = GAP(F'')$
    Compute class probabilities: $\hat{y} = \text{Softmax}(WZ + b)$
    *// Loss Computation*
    Compute weighted cross-entropy loss:

$$\mathcal{L} = -\sum_{i=1}^{K} \alpha_i \, y_i \log(\hat{y}_i)$$

    *// Backward Propagation*
    Compute gradients $\nabla_{\theta,\phi}\, \mathcal{L}$
    Update $\theta$ and $\phi$ using Adam optimizer
  **end for**
**end for**
**return** Optimized ResNet-CBAM model

**Table 5.** Training Configuration of the Proposed ResNet-CBAM Model.

| Parameter | Value |
|---|---|
| Input Image Size | $224 \times 224 \times 3$ |
| Batch Size | 32 |
| Number of Epochs | 25 |
| Optimizer | Adam |
| Initial Learning Rate | $1 \times 10^{-4}$ |
| Learning Rate Scheduler | ReduceLROnPlateau |
| Loss Function | Weighted Cross-Entropy |
| Weight Initialization | He Normal |
| Dropout Rate | 0.3 |
| Activation Function | ReLU |
| Backbone Network | ResNet-50 |
| Channel Attention Reduction Ratio ($r$) | 16 |
| Spatial Attention Kernel Size | $7 \times 7$ |
| Pooling Strategy | Global Average Pooling |
| Regularization | L2 ($1 \times 10^{-4}$) |
| Class Balancing Strategy | Augmentation + Undersampling |

 doi:10.20944/preprints202606.0372.v1



### *3.7. Baseline Models*

In the following experiments, to establish a benchmark performance for the proposed new hybrid model, we ran a series of four widely-used CNN models. These models were selected as they represent a range of different model families with respect to recurrent neural networks, efficient models for mobile devices, and accurate deep convolutional networks.

Long Short-Term Memory (LSTM)

A Long Short-Term Memory (LSTM) network was also used as a control; while in temporal applications, it can apply a series of possible spatial relationships in the features of the image following early convolutional layers Hochreiter and Schmidhuber (1997). In this case, the feature maps of a preliminary convolutional, backbone architecture, were flattened and fed into the LSTM to examine if sequence modeling would learn spatial context in waste images.

EfficientNet-B0

It was chosen due to its excellent accuracy and efficiency, achieved through the introduction of a compound scaling method, leading to an equal scaling of network depth, width and resolution Tan and Le (2019). The original model of this family EfficientNet-B0 achieves a good balance between accuracy and computational costs and is thus a good representative of the likely real-world model, which is going to be deployed on a resource constrained environment.

MobileNetV3

MobileNetV3, as a model that belongs to the group of highly efficient lightweight models that are intended to be used on mobile and edge device, is based on a hybrid of depth wise separable convolutions and a neural architecture search to achieve high performance Howard et al. (2019). Its emergence makes the need of extremely lightweight models less clear in the task of waste sorting, which is very crucial in the case of the low-power on-site sorting models.

ResNet-101

It is a key block for deep convolutional networks, as it has been able to solve the problem of degradation suffered by very deep networks (with residual connections) He et al. (2016). The ResNet-101 (101 layers) provides a benchmark of high-capacity models, that are able to learn features of the waste images, which are very complex and hierarchical, at the far ends of what can be achieved with typical, classical models.

## 4. Experimental Results and Discussion

In this section, we evaluate the proposed ResNet-CBAM framework and compare it to alternative architectures. We perform both quantitative and qualitative evaluations of performance, stability and interpretability. These analyses cover traditional performance metrics such as accuracy, precision, recall, and F1-score, as well as more sophisticated metrics such as the Receiver Operating Characteristic (ROC) area under the curve (AUC), Cohen's Kappa, Matthews Correlation Coefficient (MCC), and Jaccard Index.

Exploratory data visualization through confusion matrices and Grad-CAM heatmaps are also incorporated for insights on model performance and failures. The findings show that incorporating attention mechanisms helps with feature selectivity and boosts classification accuracy for several types of waste.

### *4.1. Experimental Setup and Software Configuration*

We used the PyTorch 2.x framework running on a Python 3.10 environment for all experiments. The proposed model and baseline models were trained on Google Colab Premium with an NVIDIA A100 Tensor Core GPU (40 GB VRAM), which facilitated fast training and convergence.

 doi:10.20944/preprints202606.0372.v1



The ResNet-CBAM model and other baseline models were developed using the `torch`, `torchvision`, and `torch.utils.data` libraries. Data augmentation and preprocessing were done using the library `torchvision.transforms` (resize, normalize, random geometric) OpenCV and Pillow were used for image processing, while Matplotlib and Seaborn were used to plot training loss, confusion matrix and Grad-CAM.

Models were trained with Adam optimizer with initial learning rate $1 \times 10^{-4}$, batch size of 32, and for 25 epochs. To bring an apples-to-apples comparison, all models were trained from scratch with the same settings.

To enhance the consistency of the experiments, we repeated training and testing using different random initializations for each model and reported the average performance. This approach ensures replicability of the results.

### *4.2. Comparative Evaluation Using Performance Metrics*

To benchmark the results obtained from the proposed ResNet-CBAM design, a comparative performance analysis was performed with four other baseline models namely, ResNet-101, EfficientNet-B0, MobileNet-V3 and LSTM. The models were evaluated using the standard metrics of classification namely Accuracy, Precision, Recall and F1-score which together give an overall picture of model reliability, sensitivity and performance across the waste types.

The results are shown in Table 6. Our proposed ResNet-CBAM model performs best across all metrics, with an accuracy of 0.9309, precision of 0.9044, recall of 0.9262 and F1-score of 0.9122. This suggests a good balance between precision and recall, which implies the inclusion of distinctive features in the model effectively discriminates among different types of waste.

**Table 6.** Overall Performance Comparison of Proposed and Baseline Models.

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| **ResNet-CBAM (Proposed)** | **0.9309** | **0.9044** | **0.9262** | **0.9122** |
| ResNet-101 | 0.9200 | 0.9100 | 0.9100 | 0.9100 |
| EfficientNet-B0 | 0.9000 | 0.9100 | 0.8900 | 0.9000 |
| MobileNet-V3 | 0.8700 | 0.8600 | 0.8600 | 0.8600 |
| LSTM | 0.8000 | 0.8000 | 0.7900 | 0.8000 |

The ResNet-101 baseline model achieves competitive results (accuracy = 0.92, F1-score = 0.91), showcasing the power of deep residual networks for capturing features. Yet, this model is deprived of a dedicated attention mechanism, so it does not explicitly emphasize the informative parts, especially in challenging visually complex scenes.

EfficientNet-B0 and MobileNet-V3 perform moderately as a result of their efficiency-accuracy trade-off. These models are highly efficient but have relatively lower recall and F1-scores, suggesting that they are less sensitive to subtle variations when distinguishing between waste types.

The LSTM model consistently performs the worst in all any metrics, suggesting that sequence-based networks are not optimal for spatially heavy image classification.

In summary, the strong performance of the proposed ResNet-CBAM model may be attributed to the use of channel and spatial attention mechanisms for feature refinement, which are instrumental in improving feature discrimination ability and inter- and intra-class variance. These findings confirm that attention-based feature refinement strategies lead to performance gains over such traditional deep learning models in the context of MSW classification.

### *4.3. Training Dynamics and Convergence Behavior*

We examined the convergence stability, generalization ability and convergence efficiency of the proposed ResNet-CBAM model over 25 training epochs to understand training dynamics. Figure 3 illustrates the training and validation accuracy and loss curves, respectively.

 doi:10.20944/preprints202606.0372.v1



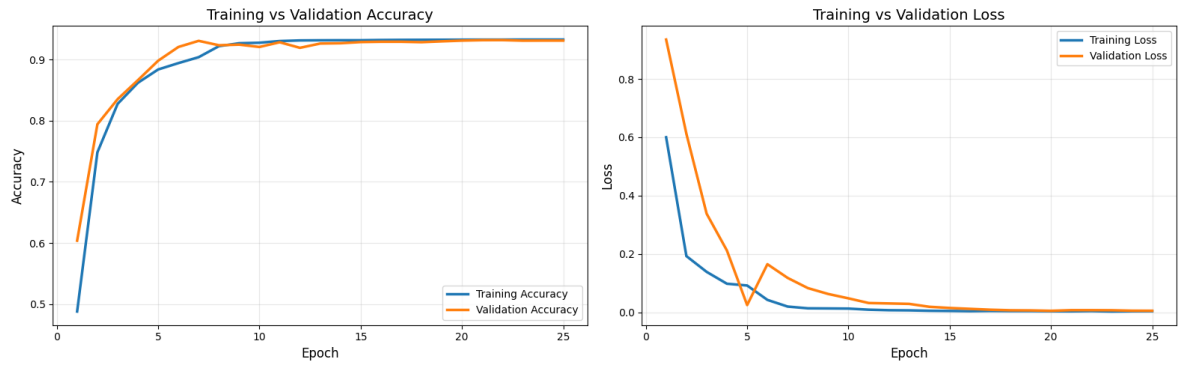


**Figure 3.** Training and validation accuracy and loss curves across 25 epochs.

As we can see (Figure 3), the model exhibits fast convergence in the early stages of training. The training accuracy rises rapidly during the first five epochs, from around 50% to more than 85%, suggesting successful initial feature extraction. Similarly, the validation accuracy increases rapidly, remaining close to the training curve.

Once the training process reaches the 10th epoch, the training and validation accuracy curves start to plateau at around 92%-94%. The narrow separation between the two curves indicates a balance between model bias and variance, and no significant overfitting.

This is also reflected in the loss curves. Training and validation losses steadily drop with each epoch, initially showing significant improvement followed by levels off towards the end of training. There are some slight variations in the validation loss in the middle epochs, which are due to noisy training (due to mini-batch-based training and data augmentation). But these fluctuations do not result in a gap between the training and validation curves.

Crucially, the absence of a noticeable discrepancy between training and validation loss suggests robust generalization. The low-convergence point of the two curves indicates that the model successfully trains to learn discriminative features from the data without overfitting the training data.

In summary, the convergence behavior indicates the combined use of CBAM attention mechanisms enhances, rather than impairs, the optimization process. The observed performance is consistent between the training and validation data, indicating that the model is suitable for real-world challenges in complex MSW classification.

### *4.4. Graphical Analysis*

To gain a full picture of classification performance, we considered several graphical and statistical measures, including Receiver Operating Characteristic (ROC) and Precision vs Recall (PR) curves, and agreement measures such as Cohen's Kappa, Matthews Correlation Coefficient (MCC) and Jaccard Index. These varying metrics offer distinct perspectives on class separability, the effects of class imbalance and the general consistency of prediction.

#### 4.4.1. Receiver Operating Characteristic (ROC) Analysis

The ROC curves for the proposed ResNet-CBAM model, shown in Figure 4, provides class-level results. The ROC analysis measures the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) at different classification thresholds, and the Area Under the Curve (AUC) provides an overall measure of the model's discriminative performance.

Our model exhibits high AUC scores across all classes, suggesting good separability. The categories with close-to-optimal separability are food waste, leaf waste and metal cans, while paper waste and plastic bottles show relatively small overlap (less separability) in features, due to their visual similarities.

Precise classification across a range of thresholds is demonstrated by the high sensitivity and low false-positive rate for the majority of categories, as shown by the massing of curves at the top-left of the ROC space.

 doi:10.20944/preprints202606.0372.v1



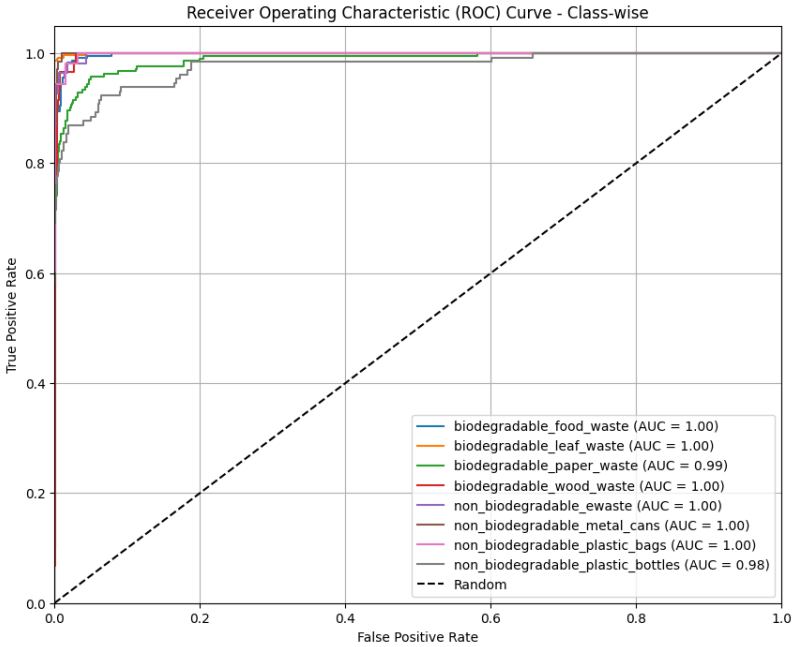


**Figure 4.** Class-wise ROC curves with corresponding AUC values.

4.4.2. Precision-Recall (PR) Analysis

The class-wise Precision-Recall curves presented in Figure 5 are useful in situations where data are imbalanced. In PR analysis, the precision (positive predictive value) and recall (sensitivity) are of particular interest.

The proposed classification model exhibits most AP values > 0.92 and approaching 1.0. Leaf waste and food waste exhibit near-perfect precision-versus-recall (PVR) relationships, whereas lower AP scores for wood waste and plastic bottles suggests somewhat harder classification tasks due to greater class overlap.

More significantly, the consistent high precision values at higher recall levels indicate robust classification with a focus on sensitivity. This is due to the attention mechanism's ability to select discriminative features.

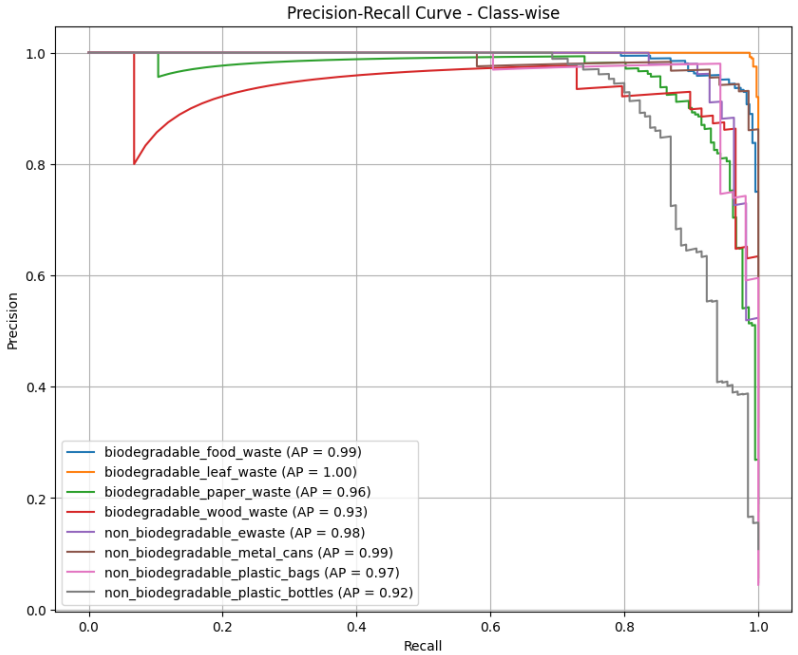


**Figure 5.** Class-wise Precision–Recall curves with Average Precision (AP) values.

 doi:10.20944/preprints202606.0372.v1



4.4.3. Agreement and Overlap Metrics

In Figure 6, we report agreement and overlap metrics, such as Jaccard Index (JI), Matthews Correlation Coefficient (MCC) and Cohen's Kappa.

Our model has a macro Jaccard Index of 0.8911, showing a large overlap of predicted and ground truth labels. The MCC of 0.9146 indicates a high correlation between predicted and true labels, using all components of the confusion matrix. Furthermore, the Cohen's Kappa of 0.9142 suggests a high level of agreement also.

These metrics indicate the model's performances are balanced between minority and majority classes, demonstrating that the model is not affected by class imbalance.

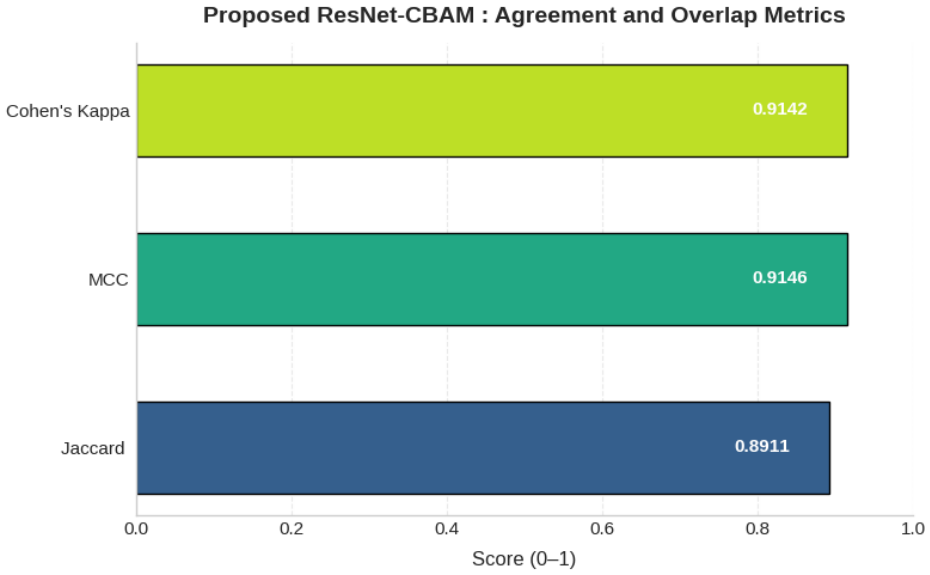


**Figure 6.** Agreement and overlap metrics (Jaccard Index, MCC, and Cohen's Kappa).

In summary, the visual assessment shows that the proposed ResNet-CBAM architecture exhibits effective discriminative ability, precision/recall balance, and inter-class agreement, making it suitable for multi-class MSW classification.

*4.5. Error Analysis*

A class-wise confusion matrix Figure 7 was used to perform a more in-depth error analysis of the proposed ResNet-CBAM model to better understand its classification performance. The confusion matrix offers a detailed breakdown of prediction results by showing the distribution of actual and predicted class labels, allowing us to detect potential misclassification patterns.

As can be seen, a large proportion of samples are predicted correctly in the diagonal, showcasing good overall predictive accuracy. Some classes, such as biodegradable leaf waste (389 correct predictions), biodegradable wood waste (57), and non-biodegradable metal cans (68) are nearly perfectly classified as shown by a few off-diagonal entries. This indicates that these classes have visual characteristics that are successfully modeled.

However, there are some misclassifications, which occur primarily between visually similar categories. For example, errors between biodegradable food waste and paper waste, wood waste, could be due to textural, color and organic matter similarities under different environmental factors. Likewise, classification confusion is seen between plastic bottles/paper bags and plastic bags/paper bags, where plastic shape transformations, transparency and brightness variations may limit the ability to discriminate between these classes.

A striking observation is that plastic bottles are misclassified as paper waste (16 times) suggesting that reflective or crushed plastic waste may be identified as paper. Further, there is some confusion

 doi:10.20944/preprints202606.0372.v1



between e-waste and wood or paper, which may imply a role of complex texture or occlusions on feature extraction.

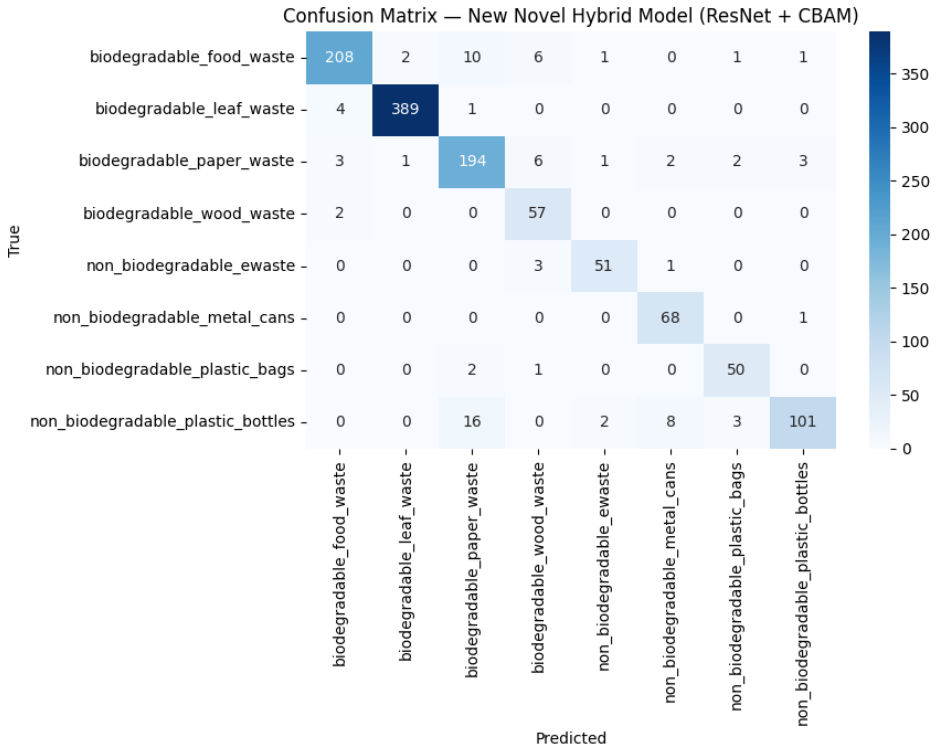


**Figure 7.** Confusion matrix of the proposed ResNet-CBAM model across eight waste categories.

Although the misclassifications may seem spread out, the overall spread is restricted and does not affect global performance scores. Furthermore, the types of errors observed are typical of real world challenges when classifying waste, including class overlap and intra-class variability.

These results emphasise that the proposed model with a enhanced architecture for better feature discrimination in ALI, but some visually complex examples remain challenging. Potential enhancements may include the use of multi-feature representations, higher resolution images or disease-specific data augmentation techniques to minimise class overlap.

In conclusion, confusion matrix analysis shows that the model has good per-class performance, with clear interpretability of remaining challenges in classification.

### *4.6. Quantitative Comparison Between Baseline and Proposed Framework*

In order to evaluate the performance of the proposed ResNet-CBAM model, a comparative analysis was performed against four baseline models (ResNet-101, EfficientNet-B0, MobileNet-V3 and LSTM). The assessment includes traditional classification metrics (Accuracy, Precision, Recall and F1-score) and additional statistical measures (Jaccard index, Cohen's Kappa coefficient and Matthews Correlation Coefficient (MCC)) to examine not only the accuracy of the models, but also robustness and generalisability of the models.

#### 4.6.1. Comparison Using Standard Performance Metrics

The results of the standard model performance metrics are shown in Figure 8. The ResNet-CBAM model we propose has the best performance across all the metrics, achieving Accuracy = 0.93, Precision = 0.91, Recall = 0.93, and F1-score = 0.92. The equal precision and recall demonstrate that the model is both highly accurate in predicting positives and highly sensitive.

 doi:10.20944/preprints202606.0372.v1



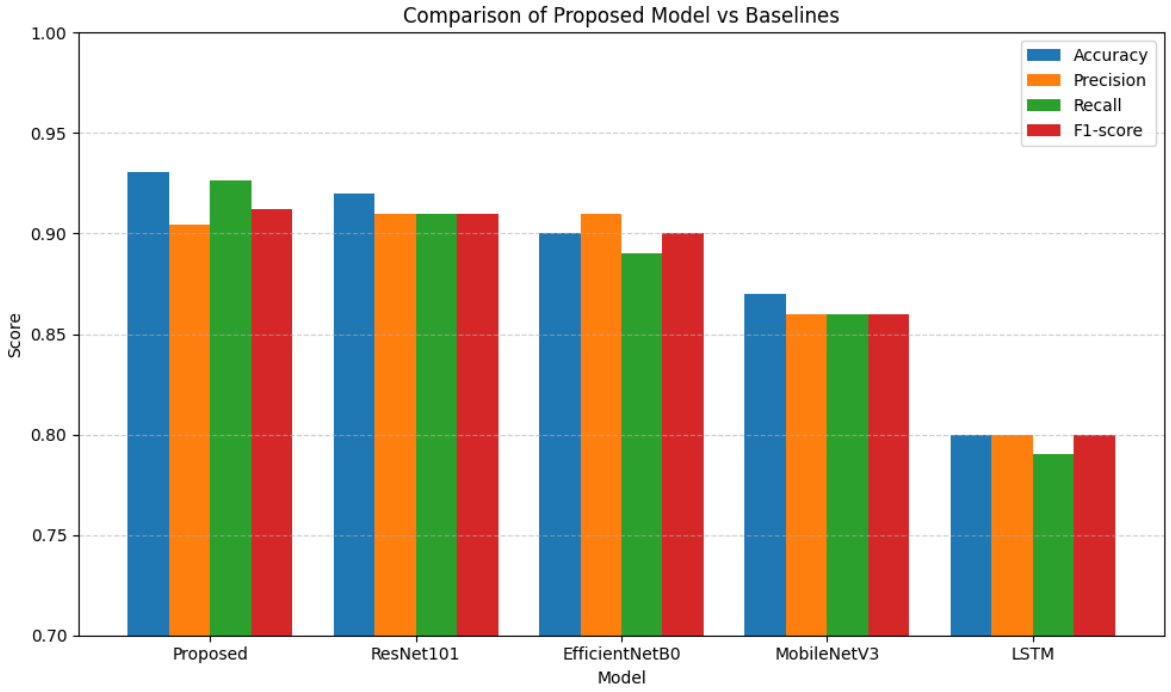


**Figure 8.** Comparison of the proposed model with baseline architectures across Accuracy, Precision, Recall, and F1-score.

The baseline model ResNet-101 also performs well, with reasonably high results. This suggests the power of deep residual learning for multi-scale feature learning. But the lack of attention mechanisms hampers its ability to concentrate on relevant areas, especially in complex visual contexts.

Modest results from EfficientNet-B0 and MobileNet-V3 reflect efficiency-performance trade-offs. Though these models are suitable for lightweight applications, their reduced recall and F1-scores imply less responsiveness to fine-scale details that discriminate between different types of waste.

The lowest metrics were found with the LSTM based model, which suggests that recurrent models may not be well-adapted to tasks with intricate spatial patterns in the image.

#### 4.6.2. Comparison Using Advanced Evaluation Metrics

The comparison of advanced statistical metrics, Jaccard Index, Cohen's Kappa and MCC is shown in Figure 9. These measures offer further information on the classification performance, agreement between classes, and overall predictor consistency.

The proposed model outperforms all others in terms of these three metrics, achieving Jaccard Index = 0.84, Cohen's Kappa = 0.92, and MCC = 0.91. The high values show high agreement between actual and predicted labels, and consistency across multiple classes.

ResNet-101, while exhibiting slightly weaker but stable performance, and EfficientNet-B0 and MobileNet-V3 have moderate agreement. The LSTM model displays the lowest performance once again, confirming that this approach is not suitable for image classification.

The high Cohen's Kappa and MCC scores of the proposed model indicate that the prediction is making use of the information learned and not blinded by class imbalance or randomness. This is likely due to channel and spatial attention blocks, which help the model prioritise valuable features while filtering out irrelevant details.

Taken together, the findings suggest that the proposed ResNet-CBAM model offers better classification performance and stability than traditional models, especially under the conditions of class imbalance and high visual similarity.

 doi:10.20944/preprints202606.0372.v1



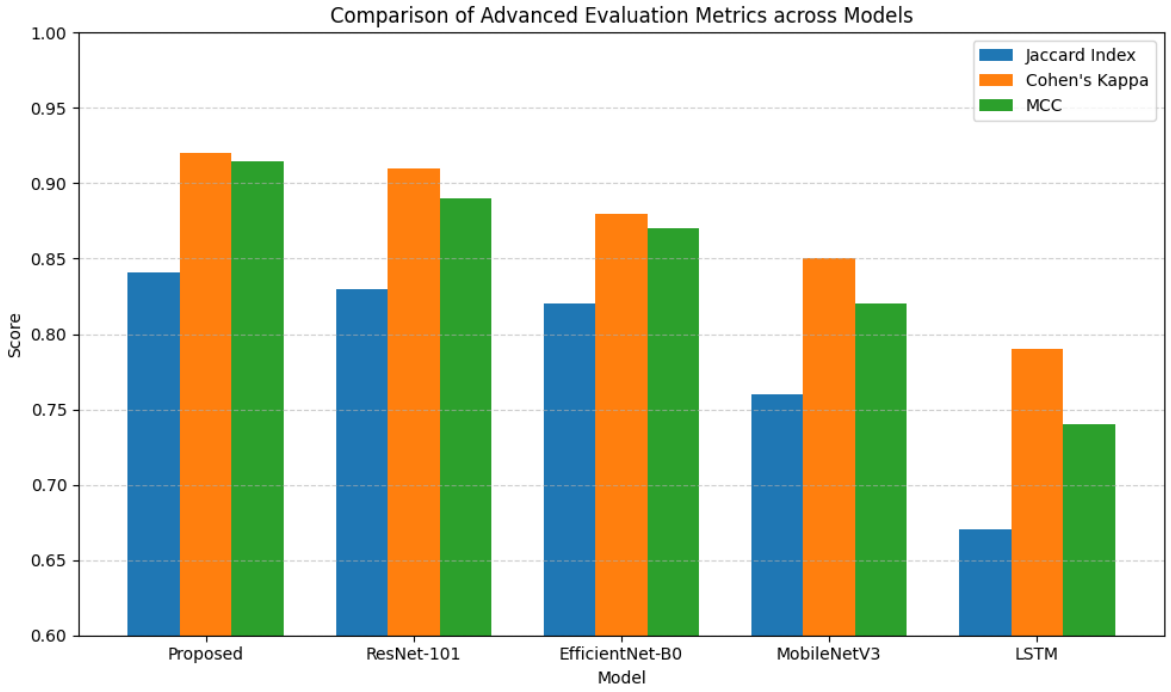


**Figure 9.** Comparison of advanced evaluation metrics (Jaccard Index, Cohen's Kappa, and MCC).

### *4.7. Ablation Study*

The effect of the components of the proposed approach was evaluated through an ablation study. It assesses the benefits of attention mechanisms on the classification task by incrementally adding channel and spatial attention to a ResNet-50 backbone. Four models were evaluated:

- **ResNet-50 (Baseline):** Original residual network without attention.
- **ResNet + Channel Attention:** Adds channel attention to highlight relevant channels.
- **ResNet + Spatial Attention:** Employs spatial attention to focus on informative locations in the image.
- **ResNet-CBAM (Ours):** Utilises channel and spatial attention together in a single module.

**Table 7.** Ablation Study on Attention Mechanisms.

| Model Variant | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| ResNet-50 (Baseline) | 0.905 | 0.892 | 0.901 | 0.896 |
| ResNet + Channel Attention | 0.918 | 0.901 | 0.915 | 0.908 |
| ResNet + Spatial Attention | 0.921 | 0.903 | 0.918 | 0.910 |
| **ResNet-CBAM (Proposed)** | **0.9309** | **0.9044** | **0.9262** | **0.9122** |

We start with a ResNet-50 model as a baseline network which shows good performance (see Table 7), and residual learning is effective in feature extraction. Adding channel attention shows significant improvement in all metrics, suggesting improved feature channel selection.

Likewise, spatial attention also leads to improved performance, as it allows the model to prioritise attention on informative regions while ignoring background noise. This is especially advantageous in waste classification where context might be an issue.

The use of the CBAM attention mechanism provides the strongest gains, leading to 2.6% higher accuracy compared to the baseline. This improvement indicates that the attention mechanisms used for channel and spatial features are complementary, and collectively enhance feature representation and feature localization.

 doi:10.20944/preprints202606.0372.v1



In summary, the ablation study shows that attention mechanisms play a positive role in improving model performance. The use of CBAM boosts discriminative feature learning and increases robustness particularly in cases where classes are similar and there exists visual variance.

### *4.8. Explainability and Feature Visualization Using Grad-CAM*

To enhance the explainability of the ResNet-CBAM model, Gradient-weighted Class Activation Mapping (Grad-CAM) is used to show the areas of the model's attention. Grad-CAM produces class-specific saliency maps that highlight the most important regions in the image for a given prediction, by calculating gradients flowing into the final convolutional layer before the global average pooling operation.

Figure 10 shows Grad-CAM results for two examples of biodegradable paper waste and biodegradable leaf waste. In the figures, the input image is displayed along with its corresponding heat map, with warmer regions representing areas that are more important for the prediction.

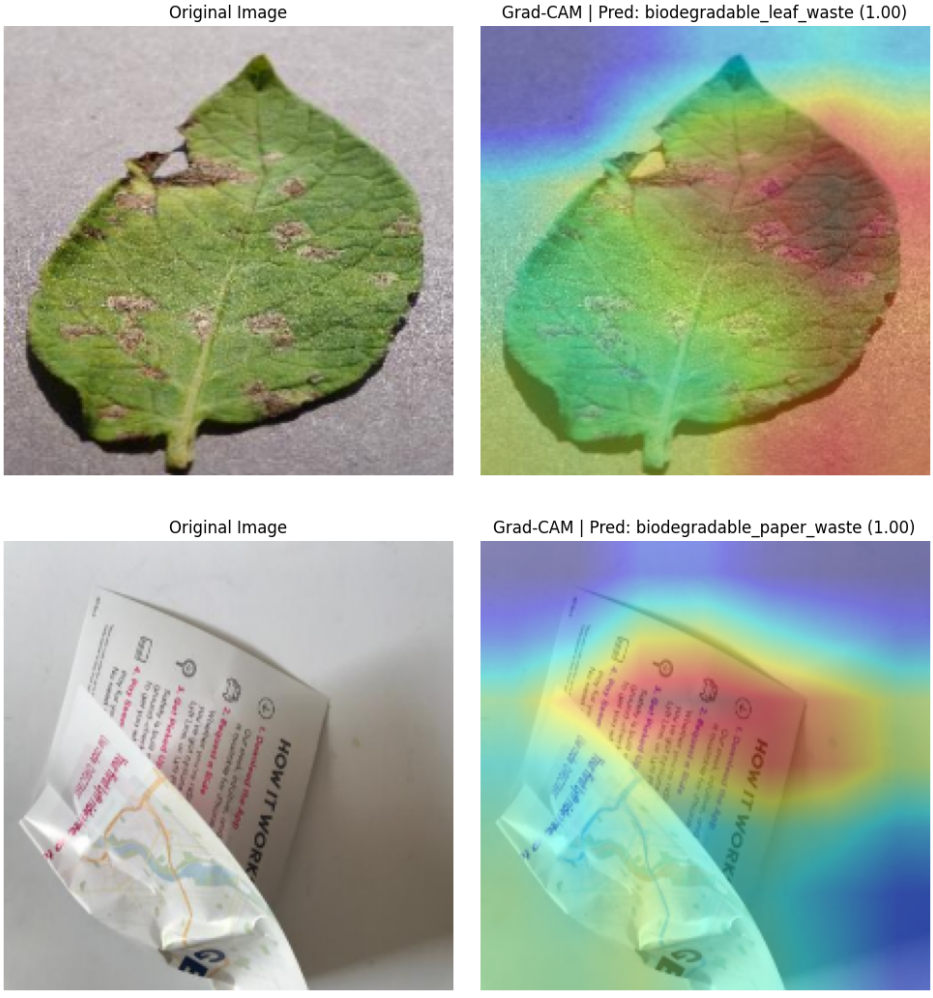


**Figure 10.** Grad-CAM visualization results for representative waste categories.

In the case of biodegradable paper waste, the heatmap mostly resembles text, folded parts, and boundaries of paper. These regions represent texture boundaries and patterns that differentiate paper from other visually similar but non-paper objects (e.g. plastic). But some blurriness is also seen in the remaining parts, which might imply that global context also partially contributes to the decision.

For the biodegradable leaf waste, the model highlights veins, irregularities and discolourations. These characteristics are potentially related to biological features of organic waste. The activation is mainly restricted to the object mask, which suggests a strong spatial localization.

 doi:10.20944/preprints202606.0372.v1



In summary, the visualizations indicate that the model focuses toward capturing relevant and class-specific features, rather than solely relying on background information. This phenomenon can be explained by the combined use of channel and spatial attention mechanisms within the CBAM module, which help the model to focus on discriminative regions of an object, suppressing the impact of background appearances.

Although the Grad-CAM visualizations offer a qualitative understanding, it's important to keep in mind that these visualizations are only approximate and may not reflect the model's decision-making entirely. However, they provide additional evidence that our architecture learns useful and meaningful representations for the task of waste classification.

These results demonstrate the interpretability of the proposed approach, which is a key to the success of future applications in real-life AI-driven waste classification systems.

### *4.9. Comparative Analysis and Discussion*

**Table 8.** Comparative Analysis of Existing Studies and the Proposed Model.

| Reference | Dataset | Model | Accuracy (%) |
|---|---|---|---|
| Alsabt et al. (2024) | World Bank Waste Dataset | SVM + RF + XGBoost + LP | 85.00 |
| Qiao (2024) | Custom Dataset (16 classes) | ResNet50 + SVM | 89.00 |
| Zhang et al. (2021) | NWNU-TRASH Dataset | DenseNet169 (TL) | 82.00 |
| Nnamoko et al. (2022) | Public Dataset (25,077 images) | Five-layer CNN | 80.88 |
| Yi and Kim (2024) | Mixed Waste Dataset | CNN + Web System | 77.62 |
| **Ours** | **Kaggle Waste Dataset** | **ResNet-CBAM** | **93.09** |

The comparative analysis in Table 8 showcases recent studies in waste classification, showcasing the variation in datasets, model approaches and performance. It's critical to note that while the studies are comparable, the differences in dataset size, number of classes, class distribution and evaluation metrics limit any direct comparison.

Weaker models based on traditional machine learning algorithms, as used by Alsabt et al. (2024), provide reasonably high accuracy (85%) but are limited in their ability to learn complicated spatial features in image data. Combined approaches incorporating deep learning and traditional classifiers, such as ResNet50 and SVM Qiao (2024), enhance feature learning but overlook feature selection in an attention-aware manner.

Deep convolutional networks (DenseNet169, Zhang et al. (2021)) and lightweight CNN variants Nnamoko et al. (2022) showcase the power of hierarchical feature representation learning but may struggle with intra-class variability and inter-class similarity issues. Likewise, practical systems like Yi and Kim (2024), designed with a focus on application, may have lower accuracy.

The ResNet-CBAM model proposed in this work reaches an accuracy of 93.09% on the Kaggle Waste Segregation dataset. Although this is a higher accuracy than other chosen studies, two studies used different set-ups from this one. However, this can be explained by the use of attention module, which enables the model to place emphasis on relevant features while ignoring the irrelevant background detail.

In conclusion, this comparative study sheds light on how attention-based frameworks can be used to enhance waste classification tasks using deep learning. But future research should incorporate cross-dataset validation and benchmarking to allow for more reliable and fair comparisons between different models.

## 5. Conclusion

This research introduced an attention-based deep learning approach for classifying municipal solid waste (MSW), namely, **ResNet-CBAM**. Introducing a ResNet-50 network with a Convolutional Block Attention Module (CBAM), it incorporates channel and spatial attentions to improve feature representations. This allows better ability to discriminate visually similar types of waste. Through

 doi:10.20944/preprints202606.0372.v1



experimentation, the proposed model exhibits an accuracy of **93.09%** and F1-score of **0.9122** on eight different types of waste. A comparative study with ResNet-101, EfficientNet-B0, MobileNet-V3 and LSTM suggests the proposed architecture offers a stable classification performance. Other value-added measures such as ROC-AUC, Precision–Recall curves, Cohen's Kappa, MCC and Jaccard Index also confirm the model's performance. The ablation study shows that the use of channel and spatial attention in conjunction improves classification performance, compared with using either attention component alone, and Grad-CAM analysis gives qualitative assurance that the model is looking at the right areas. Despite these encouraging findings, there are some caveats. First, we have tested the model on a single dataset, which may not be representative of a wide range of real-world waste scenarios. Second, while class re-sampling was used to train the model, real-world waste streams are highly variable. Third, the model uses only visual data, which may be supplemented with other sensor inputs: depth, spectral, material or any other similar information. Fourth, the model needs to be further optimised for use on low-power edge devices. Ongoing research will involve multi-dataset evaluations, and lightweight attention-based models, real-time edge deployment, and multi-modal waste classification. In conclusion, the proposed ResNet-CBAM model showcases the capabilities of attention-based deep learning for automated, interpretable, and accurate waste classification in smart waste monitoring systems.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare that they have no competing interests.

**Author Contributions:** All authors contributed equally to this work.

## References

Abogunrin-Olafisoye, O. B., & Adeyi, O. (2025). Environmental and health impacts of unsustainable waste electrical and electronic equipment recycling practices in Nigeria's informal sector. *Discover Chemistry*, *2*(1), 4.

Ahmad, G., Aleem, F. M., Alyas, T., Abbas, Q., Nawaz, W., Ghazal, T. M., Aziz, A., Aleem, S., Tabassum, N., & Ibrahim, A. M. (2025). Intelligent waste sorting for urban sustainability using deep learning. *Scientific reports*, *15*(1), 27078.

Ahmed Khan, H., Naqvi, S. S., Alharbi, A. A., Alotaibi, S., & Alkhathami, M. (2024). Enhancing trash classification in smart cities using federated deep learning. *Scientific Reports*, *14*(1), 11816.

Akhila, B., Thaile, M., Asha Kiran, M., Babu Pittala, R., & Gundalwar, P. (2025). Garbage Classification Using Convolutional Neural Networks. In *No, it has never appeared or been accepted by any conference or journal.*

Alourani, A., Ashraf, M. U., & Aloraini, M. (2025). Smart waste management and classification system using advanced IoT and AI technologies. *PeerJ Computer Science*, *11*, e2777.

Alsabt, R., Alkhaldi, W., Adenle, Y. A., & Alshuwaikhat, H. M. (2024). Optimizing waste management strategies through artificial intelligence and machine learning-An economic and environmental impact study. *Cleaner Waste Systems*, *8*, 100158.

Arishi, A. (2025). Real-time household waste detection and classification for sustainable recycling: A deep learning approach. *Sustainability*, *17*(5), 1902.

Castro-Bello, M., Roman-Padilla, D. B., Morales-Morales, C., Campos-Francisco, W., Marmolejo-Vega, C. V., Marmolejo-Duarte, C., Evangelista-Alcocer, Y., & Gutiérrez-Valencia, D. E. (2025). Convolutional neural network models in municipal solid waste classification: towards sustainable management. *Sustainability*, *17*(8), 3523.

Chen, Z., Xiao, Y., Zhou, Q., Li, Y., & Chen, B. (2024). The development of a waste management and classification system based on deep learning and Internet of Things. *Environmental Monitoring and Assessment*, *197*(1), 103.

Dipo, M. H., Farid, F. A., Mahmud, M. S. A., Momtaz, M., Rahman, S., Uddin, J., & Karim, H. A. (2025). Real-Time Waste Detection and Classification Using YOLOv12-Based Deep Learning Model. *Digital*, *5*(2), 19.

Dutt, A., & Dutt, A. (2022). *Waste segregation image dataset*. Kaggle. Available online: https://www.kaggle.com/dsv/4235079 (accessed on). https://doi.org/10.34740/KAGGLE/DSV/4235079.

 doi:10.20944/preprints202606.0372.v1



Eryeşil, Y., Kahramanli Örnek, H., & Taşdemir, Ş. (2026). Optimizing solid waste classification using deep learning and grey wolf optimizer for recycling efficiency. *International Journal of Environmental Science and Technology*, *23*(1), 44.

Fang, B., Yu, J., Chen, Z., Osman, A. I., Farghali, M., Ihara, I., Hamza, E. H., Rooney, D. W., & Yap, P.-S. (2023). Artificial intelligence for waste management in smart cities: a review. *Environmental Chemistry Letters*, *21*(4), 1959–1989.

Gunaseelan, J., Sundaram, S., & Mariyappan, B. (2023). A design and implementation using an innovative deep-learning algorithm for garbage segregation. *Sensors*, *23*(18), 7963.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., & Vasudevan, V. (2019). Proceedings of the IEEE/CVF international conference on computer vision. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 1314–1324).

Imam, S. T., & Rafizul, I. M. (2025). Occupational risks, vulnerabilities, and safety challenges among informal waste workers at the open disposal site in Khulna city. *International Journal of Hygiene and Environmental Health*, *266*, 114543.

Islam, F. (2025). Artificial intelligence-driven smart waste-to-energy networks for climate-resilient circular resource management in vulnerable megacities. *International Journal of Environment and Climate Change*, *15*(7), 381–415.

Islam, M. M., Hasan, S. M., Hossain, M. R., Uddin, M. P., & Mamun, M. A. (2025). Towards sustainable solutions: Effective waste classification framework via enhanced deep convolutional neural networks. *Plos one*, *20*(6), e0324294.

Jayaraman, V., & Lakshminarayanan, A. R. (2024). MSW-Net: a hierarchical stacking model for automated municipal solid waste classification. *Journal of the Air & Waste Management Association*, *74*(8), 569–580.

Mudemfu, M. (2023). *Intelligent solid waste classification system using deep learning* (Unpublished master's thesis). Purdue University.

Nnamoko, N., Barrowclough, J., & Procter, J. (2022). Solid waste image classification using deep convolutional neural network. *Infrastructures*, *7*(4), 47.

Oise, G., & Konyeha, S. (2025). Environmental impacts in e-waste management using deep learning. *Discover Artificial Intelligence*, *5*(1), 1–18.

Pitakaso, R., Srichok, T., Khonjun, S., Golinska-Dawson, P., Gonwirat, S., Nanthasamroeng, N., Boonmee, C., Jirasirilerd, G., & Luesak, P. (2024). Artificial Intelligence in enhancing sustainable practices for infectious municipal waste classification. *Waste Management*, *183*, 87–100.

Potharaju, S., Tambe, S. N., Tadepalli, S. K., Salvadi, S., Manjunath, T., & Srilakshmi, A. (2025). Optimizing Waste Management with Squeeze-and-Excitation and Convolutional Block Attention Integration in ResNet-Based Deep Learning Frameworks. *Journal of Artificial Intelligence and Technology*, *5*, 211–220.

Qiao, Z. (2024). Advancing Recycling Efficiency: A Comparative Analysis of Deep Learning Models in Waste Classification. *arXiv preprint arXiv:2411.02779*.

Rajakumaran, G., Usharani, S., Vincent, C., & Sujatha, M. (2023). Smart Waste Management: Waste Segregation using Machine Learning. In *Journal of physics: Conference series* (Vol. 2471, p. 012030).

Rao, R., Singh, S., Salas, M., Sarker, A., Kumar, R., Wang, Y., Lucia, L., Mittal, A., Yarbrough, J., Barlaz, M. A., et al. (2025). AI-powered municipal solid waste management: a comprehensive review from generation to utilization. *Frontiers in Energy Research*, *13*, 1670679.

Sayem, F. R., Islam, M. S. B., Naznine, M., Nashbat, M., Hasan-Zia, M., Kunju, A. K. A., Khandakar, A., Ashraf, A., Majid, M. E., Kashem, S. B. A., et al. (2025). Enhancing waste sorting and recycling efficiency: robust deep learning-based approach for classification and detection. *Neural Computing and Applications*, *37*(6), 4567–4583.

Shaibur, M. R., Siddique, A. B., Nahar, N., Al Helal, A. S., Al Maruf, M. A., Arpon, S. H., Akter, M. S., & Ambade, B. (2025). Solid waste management in an urban community of a developing country: an overview of 5Rs strategies. *World Development Sustainability*, 100254.

Sirawattananon, C., Muangnak, N., & Pukdee, W. (2021). Designing of IoT-based smart waste sorting system with image-based deep learning applications. In *2021 18th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ecti-con)* (pp. 383–387).

 doi:10.20944/preprints202606.0372.v1



Son, J., & Ahn, Y. (2025). AI-based plastic waste sorting method utilizing object detection models for enhanced classification. *Waste Management*, *193*, 273–282.

Soni, T., Gupta, D., & Uppal, M. (2024). MobileNet-based garbage classification: Enhancing recycling with machine learning. In *2024 international conference on intelligent computing and emerging communication technologies (icec)* (pp. 1–4).

Sunardi, Y. A., & Fahmi, M. (2023). Improving waste classification using convolutional neural networks: An application of machine learning for effective environmental management. *Revue D'Intelligence Artificielle*, *37*(4), 845–855.

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105–6114).

Thielmann, P., Zhou, Y., Mirbach, B., Stricker, D., & Rambach, J. (2025). A Review of Computer Vision for Industrial-Grade Waste Classification. *IEEE Access*.

Verber, D., Grneva, T., & Dugonik, J. (2026). Image-Based Waste Classification Using a Hybrid Deep Learning Architecture with Transfer Learning and Edge AI Deployment. *Mathematics*, *14*(7), 1176.

Wang, Z., Zhou, W., & Li, Y. (2024). GFN: A garbage classification fusion network incorporating multiple attention mechanisms. *Electronics*, *14*(1), 75.

Wu, L., Li, B., Du, J., & Khalingarajah, H. (2025). AI-driven predictive analytics for sustainable cities and communities: Urban waste management in Bangkok. *Big Data and Computing Visions*, *5*(3), 184–203.

Yi, C. J., & Kim, C. F. (2024). AI-Powered Waste Classification Using Convolutional Neural Networks (CNNs). *International Journal of Advanced Computer Science & Applications*, *15*(10).

Zhang, Q., Yang, Q., Zhang, X., Bao, Q., Su, J., & Liu, X. (2021). Waste image classification based on transfer learning and convolutional neural network. *Waste Management*, *135*, 150–157.