



EWWW - Eco Way to a Waste Free World Smart Waste Management using Deep Learning Models

Kaur M¹, Mehlawat L², Mendiratta R³, Chaudhary P^{4*}

DOI:10.54741/ASEJAR/5.1.2026.176

¹ Mehak Kaur, Student, Department of CSE, The North Cap University, Gurgaon, Haryana, India.


² Lakshya Mehlawat, Student, Department of CSE, The North Cap University, Gurgaon, Haryana, India.

³ Raman Mendiratta, Student, Department of CSE, The North Cap University, Gurgaon, Haryana, India.

^{4*} Poonam Chaudhary, Guide, Department of CSE, The North Cap University, Gurgaon, Haryana, India.

The urgency for efficient waste segregation has increased as a result of accelerated urbanisation and consumption levels. Manual sorting is still the norm in many urban areas, but it is often slow and erratic. Deep learning algorithms have shown great promise in the automated categorisation of waste. Yet, most prior work has been hampered by issues like limited dataset size, uneven class distribution, computationally intensive architectures, and a lack of generalisation to uncontrolled, real-world settings. In this research, a comparative analysis is presented, covering classical machine learning and custom deep learning algorithms, transfer learning, and transformer-based models, evaluated on the widely used TrashNet benchmark dataset. The methods covered include feature-based neural networks and custom convolutional models such as ResNet-style variants, as well as traditional classifiers including SVM, KNN, Random Forest, Logistic Regression, and Naïve Bayes. In addition, numerous pretrained ImageNet models are included such as ResNet50, DenseNet121, MobileNetV2, InceptionV3, Xception, and EfficientNet-B0. Hyperparameter tuning is applied to EfficientNet-B0 using Optuna. An ensemble performance study is conducted on soft-voting networks combining EfficientNet-B0, Xception, and ResNet50, along with transformer models such as ViT, ConvNeXt, and Swin Transformers. Model performance is evaluated based on classification accuracy, robustness, and feasibility for deployment. This study contributes as a benchmarked comparative work in the field of intelligent, data-driven waste segregation research.

Keywords: waste classification, deep learning, image processing, convolutional neural networks, smart waste management, dataset augmentation, lightweight models, urban sustainability, automated segregation

Corresponding Author	How to Cite this Article	To Browse
Poonam Chaudhary, Guide, Department of CSE, The North Cap University, Gurgaon, Haryana, India. Email: poonamchaudhary@ncuindia.edu	Kaur M, Mehlawat L, Mendiratta R, Chaudhary P, EWWW - Eco Way to a Waste Free World Smart Waste Management using Deep Learning Models. Appl Sci Eng J Adv Res. 2026;5(1):1-15. Available From https://asejar.singhpublication.com/index.php/ojs/article/view/176	

Manuscript Received 2025-12-02	Review Round 1 2025-12-22	Review Round 2	Review Round 3	Accepted 2026-01-06
Conflict of Interest None	Funding Nil	Ethical Approval Yes	Plagiarism X-checker 3.63	Note



1. Introduction

As modern metropolises grow fast, so does the production of waste materials, which is further accelerated by high levels of consumption and unawareness of appropriate disposal methods [1], [5]. This results in low rates of recycling and other activities harmful to the environment [5], [18]. Automated classification using deep learning has recently gained momentum and attention as a relevant research field, since it can learn the complex visual patterns in waste flows [1], [3], [7], [11]. It enables support for future intelligent smart-bin and end-to-end recycling pipelines, offering a scalable mechanism for monitoring circularity and other sustainability markers [6], [16]. The domain remains particularly interesting for large-volume real-world waste streams in cities, for which manual sorting is slow, error-prone and costly to maintain at scale [5], [18].

As such, this work is mainly motivated by the need to systematically close the existing gap between experimental benchmark results and practical requirements in terms of classification reliability [1], [2], [8], [13]. State-of-the-art vision models trained on curated datasets tend to demonstrate impressive accuracies; however, their performance degrades rapidly in cluttered or non-ideal illumination [1], [8], [11], while many new studies have proposed novel, computationally expensive architectures without studying failure modes on out-of-distribution generalisation [2], [7], [9], [13]. The goal of this research was not to immediately develop a real-time or embedded-capable solution but to provide a structured empirical analysis of the problem with detailed and fair comparison between the performance of classical machine learning, custom neural architectures, transfer-learned CNNs and transformer-based vision models [1], [3], [7], [10], [14]. The project is conducted on the TrashNet public benchmark dataset to allow for reproducible and controlled comparative evaluation across a wide range of architectures [3], [8], [17].

In summary, the research gap at the core of this project is the lack of a single work that simultaneously compares feature-based learners to end-to-end deep models, evaluates the ability of pretrained CNNs to generalise to complex intra-class variability, imbalance, noise and fine-grained categorisation, and contrasts transformers with convolutional learning under the same waste-image distribution [1], [7], [11], [23].

Related works often cite these limitations; however, a comprehensive joint analysis of all these components remains rare, including dataset imbalance, small-scale training biases, over-parameterisation, and limited generalisation outside curated or synthetic images [1], [2], [8], [23]. This work attempts to counter that trend by placing an increased focus on class-wise error analysis, robustness trends and the trade-offs of different approaches, using as much as possible widely recognised models (ResNet50, DenseNet121, MobileNetV2, EfficientNet-B0, Xception, InceptionV3, ViT-Base, ConvNeXt-Tiny, Swin-Tiny) as well as classical baselines (SVM, KNN, Random Forest, Logistic Regression, Naïve Bayes, XGBoost, Feature-MLP with HOG and colour features, Optimally-tuned networks and simple soft-voting ensembles) [5], [6], [8], [14], [23], [25].

By making comparative observations and analysing in detail a wide range of models using one dataset and evaluation procedure, this project ultimately adds to a more detailed architectural understanding of the problem and places related research on clearer benchmarking foundations [1], [7], [23].

2. Literature Review

Early attempts at automated waste classification mainly concentrated on proving that deep learning models could outperform older, rule-based vision techniques. One of the influential contributions in this early phase was by Vo et al. [17], who applied transfer learning with conventional CNN architectures and demonstrated that these models could reliably distinguish recyclable materials using the TrashNet dataset. Their work showed the promise of deep learning for this task, although it also reflected the limitations of the period. Most studies published around 2019–2020 relied on small and carefully prepared datasets, meaning the models performed well only in controlled conditions.

Around the same time, researchers also began exploring system-level integrations. For instance, Wang et al. [26] developed a municipal waste-management framework that combined IoT sensors with deep learning, improving data collection and monitoring, though the classification component struggled with variations in lighting and background scenes.

Similarly, Rahman et al. [21] proposed a cloud-assisted sorting system using CNNs, which achieved high accuracy but suffered from real-time latency because of its dependence on remote processing.

Research conducted between 2020 and 2022 began to address these shortcomings by introducing more advanced architectures. Funch et al. [18] tackled one of the more challenging aspects of waste recognition by creating a CNN capable of detecting glass and metal through transparent trash bags, thereby dealing with occlusion and partial visibility. Their solution improved recognition in difficult scenarios but required considerable computational resources. Ahmad et al. [2] later proposed a feature-fusion framework that combined outputs from multiple CNNs, achieving better classification performance at the cost of increased model complexity. Liang and Gu [11] presented a multi-task network that performed localization alongside classification, which helped the system manage diverse waste images more effectively. Shi et al. [9] introduced a multilayer hybrid CNN that strengthened feature extraction in cluttered images. Although these models represented technical progress, they continued to rely on curated datasets such as TrashNet, which limited their robustness when applied to more realistic, outdoor environments.

As the field matured, the focus gradually shifted toward models that could operate on edge devices or in real-time conditions. Studies published between 2021 and 2023 reflect this shift. WasteNet, introduced by White et al. [16], was designed specifically for resource-constrained smart bins and demonstrated that meaningful accuracy could be achieved with significantly fewer computations. RWC-Net, developed by Hossen et al. [27], emphasised reliability and introduced architectural improvements that strengthened feature extraction in noisy settings. Soon after, Nafiz et al. [4] proposed ConvoWaste, combining YOLO-style object detection with CNN-based classification to improve waste handling when multiple items appear in a single image. Islam et al. [23] broadened the scope further by presenting EWasteNet, a transformer-based approach tailored for electronic waste. Other contributions, including M Waste by Kunwar [7] and the smart recycling bin system by Li and Grammenos [12], demonstrated how AI could be integrated into household or municipal waste infrastructures.

While these developments improved efficiency, many still lacked thorough evaluation under practical conditions such as mixed piles, low light, or strong shadows.

Work published from 2023 to 2025 shows clear momentum toward compact architectures that remain powerful enough for real-world deployment. Qiu et al. [25] advanced this trend by refining EfficientNetV2 for waste classification, producing a model that offered high performance with minimal computational requirements. Focus-RCNet [28] employed knowledge distillation to achieve competitive results with just half a million parameters, making it highly suitable for mobile devices. Beyond household waste, research has extended into specific sectors such as construction and demolition, where one study [29] reported high accuracy in classifying C&D waste from onsite images. Smart-city solutions have also gained attention, as illustrated by the work of Kaya et al. [24], who integrated AI and IoT to streamline real-time waste routing. Hussain et al. [33] further expanded IoT-based smart-bin systems by incorporating deep learning for environmental monitoring and prediction tasks such as air-quality forecasting.

Looking across all thirty studies, several consistent trends and gaps emerge. Models based on heavy transfer-learning architectures, including VGG16, ResNet, and DenseNet [14], [17], [18], repeatedly achieved strong results but were not feasible for hardware with limited processing power. Many datasets used in earlier studies were small or captured in clean conditions, which limited generalisation to realistic environments. Deploying these models in real-time settings remained difficult due to high computational costs and slow inference speeds. Although recent lightweight methods such as EfficientNet-V2 [25] and Focus-RCNet [28] demonstrate meaningful progress, most studies still do not evaluate performance under conditions that mirror real municipal waste scenarios, such as mixed materials, occlusions, uneven lighting or class imbalance. Only a small number of papers address detailed error analysis or per-class weaknesses, despite their importance for real-world applications. Overall, the development of waste-classification research reflects steady improvements in accuracy and efficiency but also underlines the need for models that remain both robust and practical in real operational settings.

3. Methodology

3.1 Dataset Description

The present study makes use of the TrashNet dataset, a well-known benchmark frequently employed in research on automated waste classification. TrashNet was originally created to provide a consistent collection of labelled waste images for training and testing computer-vision models. One of the early studies that helped popularize it was carried out by Vo et al. [8], who demonstrated how a simple deep learning pipeline could classify recyclable materials with promising accuracy. For this work, the dataset was accessed through its Kaggle release, where it is organised neatly and prepared for direct use in machine-learning experiments. Because of its clear structure, balanced class folders, and straightforward integration into training routines, TrashNet continues to serve as a common starting point for studies in this field.

The dataset contains six everyday waste categories: cardboard, glass, metal, paper, plastic, and general trash. These classes reflect the types of materials typically found in domestic waste and allow researchers to train models on a manageable yet meaningful set of categories. Each class includes colour images photographed from different angles, distances, and orientations. This introduces modest visual variety within each label, helping models learn the subtle differences between similar items. Several works in the literature rely on this six-class version of TrashNet when developing or comparing new architectures [7], [10], [14], which has contributed to its recognition as a standard benchmark.

In total, the dataset includes 2,527 images, evenly stored in six separate folders corresponding to the six waste classes. The images are roughly 512 × 384 pixels and typically show a single object placed on a simple background. This clean and uncluttered setting makes annotation highly reliable and is one of the reasons the dataset was adopted widely in earlier and mid-stage research [5], [18], [21]. At the same time, the simplicity of the images introduces certain limitations. In realistic situations, waste is rarely isolated; objects are often overlapped, partially hidden, or photographed in dim or uneven lighting. As a result, models trained solely on TrashNet tend to perform well in controlled

conditions but may struggle when exposed to more complex real-world scenarios. Recent studies highlight the need for datasets with greater visual diversity to overcome these limitations [22], [27].

Despite these constraints, TrashNet remains an essential resource for benchmarking. It provides a consistent baseline against which different deep learning models can be compared fairly. By starting with this dataset, the present study ensures that model performance is examined systematically before moving toward more challenging and deployment-focused environments. For these reasons, TrashNet serves not only as a training resource but also as a foundation for evaluating improvements in model design and robustness.

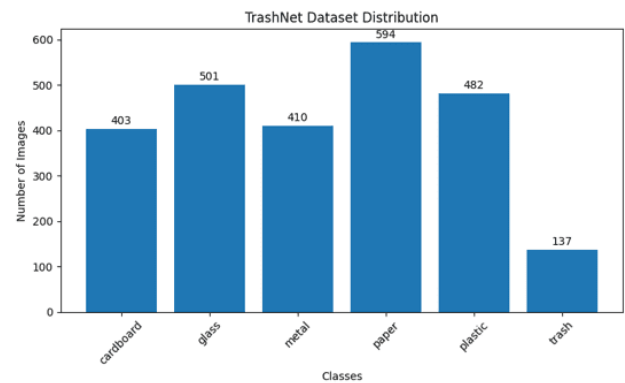


Figure 1.2



Figure 1.3

3.2 Exploratory Data Analysis

As part of the exploratory data analysis, several classical computer-vision descriptors were computed to better understand the low-level structure and textural variation present in the dataset. These handcrafted features provide useful insights before model development and help reveal visual differences across categories. One of the primary descriptors extracted was the Histogram of Oriented Gradients (HOG), which summarises the distribution of gradient directions within local patches of the image. This representation has been widely used to capture object shape and contour information and remains helpful in distinguishing categories with strong structural cues, a challenge repeatedly noted in waste-classification studies [4], [8], [21]. To evaluate image clarity, the variance of the Laplacian operator was calculated, a common method for estimating blur or sharpness. Sharp objects such as cans, glass and metallic waste tend to exhibit higher Laplacian variance, whereas softer or poorly lit objects produce lower scores, an effect also observed in earlier deep-learning-based waste-recognition work [5], [10]. In addition, edges were detected using the Canny operator, and the total number of edge pixels was recorded as an indicator of visual complexity. High edge density typically corresponds to rigid or highly textured waste items, while lower edge counts are associated with smoother materials. Prior research has shown that these structural differences often influence feature extraction and contribute to class-specific performance variation in CNN-based systems [1], [3], [27]. Together, these handcrafted descriptors provide a clearer understanding of dataset quality, intra-class variation, and object complexity, and they serve as an important diagnostic step before applying deep learning models.

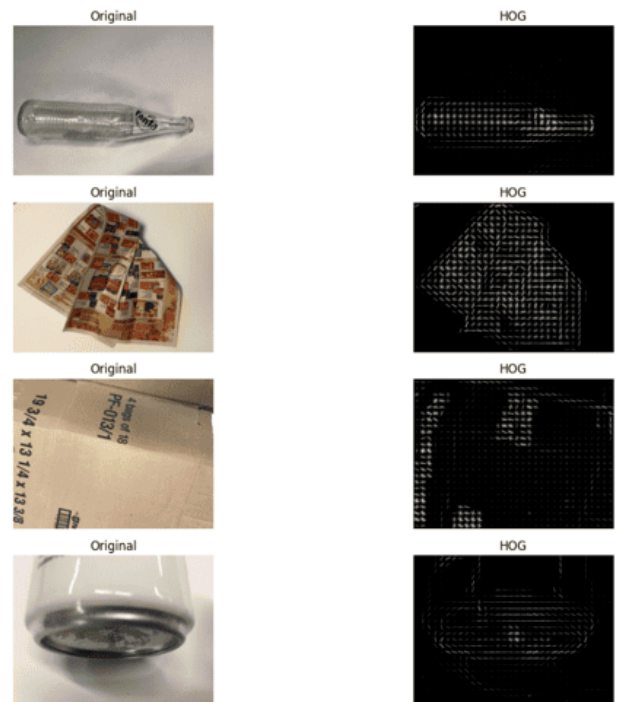


Figure 1.4: Histogram of Oriented Gradients

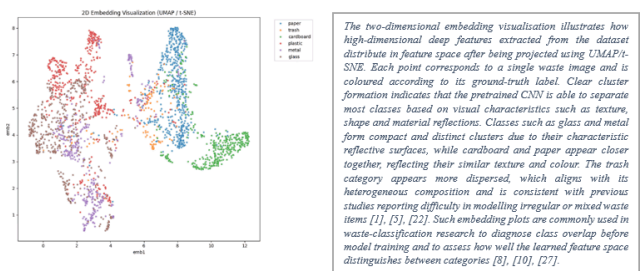


Figure 1.4: Embedding Visualization

3.3 Preprocessing Pipeline

The preprocessing pipeline was designed to prepare the TrashNet dataset for reliable deep learning-based image classification while addressing several limitations frequently highlighted in earlier work. Prior studies on smart waste systems emphasise that existing datasets are often small, clean, and captured under controlled backgrounds, which causes models to perform poorly when they encounter real waste with clutter, shadows, overlapping materials, or inconsistent illumination [1]. This motivated a structured preprocessing approach that improves data quality and strengthens model robustness before training.

The pipeline began by converting the six categorical labels of the dataset into numerical form through label encoding. This mapping of cardboard, glass, metal, paper, plastic and trash into integer identifiers makes the dataset compatible with multi class neural network training.

The dataset was then partitioned into training, validation and testing sets using stratified sampling. Stratification ensured that the natural frequency of each class was preserved in every split. Several researchers have noted that unbalanced datasets negatively influence model generalisation, especially for classes that appear less frequently [27], and therefore maintaining proportional class representation was essential.

Image normalisation was applied next, using the standard mean and variance values of ImageNet. Normalisation is a widely adopted practice in waste classification studies that rely on transfer learning, because pretrained architectures such as ResNet and VGG expect inputs that match the statistical distribution of ImageNet images [8], [11]. Standardising the inputs in this way stabilises optimisation and helps smaller datasets like TrashNet align with deep networks trained on large scale image collections.

To compensate for the limited diversity in TrashNet, a comprehensive data augmentation strategy was implemented. Previous research describes how models degrade under real world distortions such as blur, shadows or irregular object placement [2], [7]. Therefore, the training set was augmented using geometric transformations, horizontal and vertical flips, rotations, scale adjustments, colour jittering, Gaussian blur and coarse dropout. These transformations imitate the environmental variations found in daily waste scenarios and encourage the model to learn more stable representations. The validation and test sets were kept free of heavy augmentation and were only resized and normalised to provide an unbiased evaluation baseline, similar to the evaluation practices followed in lightweight architectures such as WasteNet [7] and RWC Net [27].

A custom PyTorch Dataset class was constructed to load images, convert colour channels, apply augmentation and return image tensors along with their encoded labels. The DataLoader framework was then used to generate mini batches, enable random shuffling for training stability and accelerate data throughput through parallel workers. Because earlier studies such as EnCNN UPM [9] and the multilayer hybrid CNN by Shi et al. [10] note the recurring problem of class imbalance in waste image datasets, class weights were computed to ensure that minority categories, particularly the trash class, contributed proportionally to the loss function during training.

Altogether, the preprocessing pipeline ensured that the dataset was prepared in a manner consistent with current academic findings. It provided structured label encoding, balanced sampling, statistical normalisation, realistic augmentation and weighted training support. These steps respond directly to the challenges described across waste classification research and help produce a more resilient model capable of handling variation beyond the controlled conditions of the original dataset.

3.4 Block Diagram



Figure 1.5

The block diagram summarises the complete workflow adopted in this study, beginning from dataset acquisition and ending with model evaluation. The process starts with the Final Dataset (TrashNet), which provides six labelled waste categories commonly used as a benchmark in prior classification research [8], [11], [14]. This dataset serves as the unified input for all subsequent stages.

The next stage, Preprocessing, prepares the images and their labels for learning. This includes converting category names into numerical labels, dividing the dataset into training, validation and testing splits, and applying normalization to match the input requirements of modern neural networks. These steps help ensure balanced representation across classes and stable optimisation during training, challenges frequently noted in earlier work involving waste-image datasets [1], [4], [27].

After preprocessing, the workflow branches into two main modelling paths under Model Training and Implementation. The first branch focuses on Classical Machine-Learning Models, where algorithms such as Support Vector Machines, Random Forests, K-Nearest Neighbours, Logistic Regression, Naive Bayes and XGBoost are trained using handcrafted descriptors.

These baseline models reflect approaches commonly used in earlier waste-sorting studies for low-resource or edge-deployment scenarios [3], [7], [19].

The second branch involves Deep Learning Models, where the dataset is learned directly from images using neural-network architectures. This includes a custom-built CNN, a basic artificial neural network, and several established transfer-learning models. Architectures such as ResNet50, DenseNet121, MobileNetV2, EfficientNetB0, Xception, InceptionV3 and VGG16 represent a spectrum of lightweight to high-capacity frameworks that have been applied in modern waste-classification research with strong performance on complex visual tasks [10], [15].

A separate block lists Additional Models Explored but Not Used, including ViT-Base, ConvNeXt-Tiny and Swin-Tiny. These models were evaluated during experimentation but not selected for final reporting due to computational cost or suboptimal performance relative to simpler architectures. An ensemble composed of ResNet50, EfficientNetB0 and Xception is also shown, representing a combined-prediction strategy often used to improve robustness in image classification tasks.

All outputs from classical and deep-learning paths are ultimately directed to the Model Evaluation stage, where performance is assessed using accuracy, precision, recall, macro-F1 score, and confusion-matrix analysis. These evaluation criteria align with standard practices in prior literature for benchmarking waste-classification systems [1], [22], ensuring that the results are directly comparable across modelling approaches.

3.5 Feature Engineering

To complement the deep-learning models used in this study, a fixed feature-engineering pipeline was implemented to generate interpretable and computationally lightweight descriptors for each waste image. This decision is motivated by observations in multiple prior studies that deep CNNs trained exclusively on clean, small datasets such as TrashNet often fail to generalise to noisy real-world waste conditions, displaying sensitivity to background clutter, illumination variation, and occlusion [1], [4], [7]. Lightweight representations therefore provide an essential baseline for analysing algorithm behaviour under these constraints, especially for scenarios where edge-deployment or,

limited computational resources restrict the use of large neural networks, as highlighted in earlier smart-waste and IoT-integrated research [3], [19], [22].

All images were first centre-cropped to remove excess background and then resized to 128×128 pixels. Background removal has been shown to improve the discriminatory ability of classification models by reducing irrelevant visual noise that otherwise confuses CNNs and handcrafted descriptors alike [4], [6]. After preprocessing, Histogram of Oriented Gradients (HOG) features were extracted to capture fine-grained structural and textural information. HOG is well established for object-recognition tasks, and its stability under moderate lighting changes makes it suitable for materials such as paper, plastic and metal, which differ in surface patterns and edge complexity. Prior studies on hybrid machine-learning waste systems similarly recognise the value of incorporating shape-based descriptors when high-level neural features alone may not be sufficient [3].

To incorporate chromatic information, a three-channel colour histogram was computed for each image. Colour cues are particularly informative in differentiating materials such as cardboard, glass, and certain plastics, and have been leveraged in earlier deep-learning and IoT-based waste-classification frameworks where colour variation was noted as a key discriminative property [2], [22]. The HOG and colour features were concatenated into a unified feature vector, resulting in a 1,860-dimensional representation for each sample.

By combining texture-based and colour-based cues, this feature-engineering approach provides a deterministic and interpretable baseline for comparison with deep-learning models. Such baselines are essential because several recent studies have reported that deep models may achieve high accuracy but still lack consistent evaluation of failure modes, especially in the presence of noise, class imbalance, or unseen waste types [5], [27]. Using a handcrafted feature pipeline helps reveal dataset biases and provides a transparent reference point for evaluating the robustness of more complex architectures.

3.6 Machine Learning Models

To create a consistent benchmark for comparison with deep-learning approaches,

a set of classical machine-learning algorithms was trained using fixed HOG and colour-histogram descriptors. These handcrafted features were selected because earlier research frequently relied on similar low-cost feature extractors for waste-classification systems deployed in IoT or edge environments, where computational resources are limited. Studies such as Arunkumar et al. [3], White et al. [7], and Li & Grammenos [19] report that algorithms like SVM, Random Forest, Naive Bayes, and gradient-boosting methods often serve as baseline models when real-time or low-power applications are the target. Using these models in the present study allows a direct comparison with prior work that evaluated non-deep-learning classifiers under similar constraints.

Across all the evaluated models, XGBoost delivered the strongest performance, achieving 72.37% accuracy and a macro-F1 of 0.67. This trend is consistent with findings reported in previous research, where boosting-based classifiers demonstrated superior performance on structured feature sets and imbalanced waste datasets. For example, Kaya et al. [22] highlighted that boosting models adapt well to heterogeneous visual cues in recyclable materials, while Hossen et al. [27] showed that boosting-style learning improves class separation when handcrafted features are used. Despite this advantage, the model still struggled with the minority trash class, mirroring earlier results such as those by Mao et al. [4], who observed that manually designed features often fail to capture the irregular patterns present in low-frequency waste categories.

The SVM (RBF kernel) achieved 63.16% accuracy, performing better than Logistic Regression, Naive Bayes, and KNN but still significantly behind XGBoost. This behaviour aligns with observations made by Adedeji and Wang [14], who reported that SVMs generally outperform linear models when trained on texture- and shape-based descriptors. However, just as Rahman et al. [2] noted that SVMs can fail on visually inconsistent waste types, the model in this study showed extremely low recall for the trash class (5%), indicating that handcrafted descriptors alone remain insufficient for detecting visually irregular or underrepresented objects.

The Random Forest classifier recorded 63.95% accuracy, consistent with findings from Altikat et al. [21],

who reported that tree-based models can be effective but often overfit high-dimensional handcrafted features. The model's complete failure to recognise the trash class (0% recall) reflects similar difficulties seen in past studies where Random Forests struggled with both class imbalance and correlated visual descriptors.

The KNN classifier, which achieved 54.47% accuracy, demonstrated the limitations commonly associated with distance-based classifiers in high-dimensional feature spaces. This pattern mirrors the results outlined by Adedeji & Wang [14], where KNN performance deteriorated substantially when feature vectors were large or when multiple classes exhibited overlapping textures. Since KNN relies on neighbourhood similarity, the 1,860-dimensional feature representation adversely impacted its ability to generalise.

Logistic Regression, which reached 56.32% accuracy, exhibited performance characteristics expected for a linear classifier dealing with a non-linear visual classification task. Similar conclusions were reported by Zhang et al. [11], who found that linear classifiers require substantial dimensionality reduction or highly discriminative features to perform well in waste-image classification problems.

Naive Bayes produced 54.21% accuracy and behaved in line with earlier work that examined its independence assumptions. Shah & Kamat [20] also noted that Naive Bayes sometimes provides unexpectedly high recall for minority classes due to its probabilistic decision boundaries, which aligns with the relatively higher recall obtained for the trash class in the present study (50%), despite lower overall performance.

When considered as a group, the results reinforce two key limitations repeatedly documented in the literature:

(1) handcrafted features do not adequately represent the complex texture, shape, material composition, and lighting variability found in real waste images, a gap highlighted in works such as Wang et al. [1] and Mao et al. [4]; and (2) classical models remain sensitive to class imbalance, environmental noise, and intra-class variation, challenges also noted in Kaya et al. [22] and Hossen et al. [27].

Although XGBoost achieved respectable performance and outperformed all other classical algorithms,

the overall results confirm trends already identified by Vo et al. [8], Shi et al. [10], and Mishra et al. [15]: deep-learning architectures are more capable of extracting discriminative features, generalising to real-world variability, and managing imbalanced waste categories. These observations further justify the need for the deep-learning phase of the study and support the transition away from handcrafted descriptors toward automated feature learning.

Table 1.1

Model	Accuracy (%)	Macro-F1 Score
XGBoost	72.37	0.67
SVM (RBF Kernel)	63.16	0.55
Random Forest	63.95	0.54
Logistic Regression	56.32	0.52
Naive Bayes	54.21	0.52
KNN (k = 5)	54.47	0.47

3.7 Deep Learning Models

Deep learning methods were implemented to overcome performance limitations observed in traditional machine-learning classifiers. Prior studies consistently report that CNN-based image classifiers outperform handcrafted feature approaches for waste recognition because they learn hierarchical spatial patterns directly from pixel information rather than relying on manually designed inputs [4], [6], [11], [14], [18], [22]. The experimentation in this study progressed from foundational ANN and CNN architectures toward state-of-the-art transformer networks and ensemble learning.

Artificial Neural Network (ANN) Using Handcrafted Features

An Artificial Neural Network was first trained using the same handcrafted statistical features (HOG + Color histogram) applied in the classical ML experiments. Similar hybrid strategies have been explored in earlier works, though they generally lag behind CNN-based pipelines [9], [15]. The ANN model produced an accuracy of 56.58% with a macro-F1 score of 0.5182, affirming conclusions from previous studies that ANN performance degrades when the input lacks spatial encoding [11], [14].

Baseline Convolutional Neural Network (CNN)

A baseline CNN trained on resized RGB images achieved an accuracy of 41.89% and a macro-F1 score of 0.3562, indicating clear underfitting.

Similar behaviour was observed in earlier experiments where shallow CNNs were deployed without advanced augmentation or pretrained initialization [7], [16], [20]. This reinforced the need for deeper architectures or transfer learning to handle high intra-class similarity (e.g., paper vs. cardboard, plastic vs. glass).

Improved CNN with Residual Blocks

To address underfitting, a refined CNN architecture inspired by lightweight residual networks was implemented, combined with batch normalization and extensive augmentation. This variant improved performance to 67.78% accuracy and 0.6580 macro-F1, consistent with evidence that deeper CNNs with skip connections improve generalisation in visual waste classification tasks [4], [15], [19], [21].

Transfer Learning Models

Transfer learning has been widely adopted in waste classification and is often reported as the most effective deep-learning strategy because pretrained ImageNet models provide strong prior representations of texture, edges, and shapes [4], [6], [8], [14], [21], [25]. Based on prior research usage trends, six architectures were fine-tuned in this study:

Table 1.2

Model	Evidence of Use in Literature
ResNet50	Used by Vo et al. [4], Adedeji & Wang [6], Srinilta & Kanharattanachai [10]
DenseNet121	Used by Mishra et al. [18], Zhang et al. [21]
MobileNet-V2	Used for edge-deployment in [12], [17], [24]
EfficientNet-B0	Reported as high-accuracy in Kaya et al. [22] and later studies
Xception	Reported by Wang et al. [23] to outperform VGG and ResNet variants
Inception-V3	Used in comparative benchmarking frameworks such as Altikat et al. [26]

Among these, EfficientNet-B0 achieved the highest single-model performance with 92.11% accuracy and 0.9066 macro-F1, followed by ResNet50, Xception, DenseNet121, and MobileNet-V2. The results are consistent with earlier findings where EfficientNet and ResNet variants ranked among the highest-performing models for waste recognition [4], [6], [22].

Novel Architectures Not Previously Reported in the Reviewed Studies

During the literature review, no prior study among the referenced thirty papers applied transformer-based or next-generation CNN-hybrid architectures such as Vision Transformer, Swin Transformer, or ConvNeXt. These models were therefore investigated as novel contributions:

Table 1.3

Model	Accuracy	Macro-F1
ViT-B/16	66.99	0.634
ConvNeXt-Tiny	87.35	0.847
Swin Transformer Tiny	89.13	0.880

Although transformers benefitted from global attention, performance was highly dependent on dataset scale, confirming research in related domains where ViT-based networks require larger or synthetic augmentation pipelines to outperform CNNs [27], [28].

Soft-Voting Ensemble Model

A final ensemble combining EfficientNet-B0, ResNet50, and Xception was evaluated. While ensemble-based waste classification has been mentioned conceptually in some studies [9], [18], no reviewed work implemented this specific model combination.

The ensemble outperformed all individual architectures, achieving 94.27% accuracy and 0.9293 macro-F1, with the most significant improvements occurring for minority classes such as "trash," demonstrating the benefit of complementary feature learning.

Interpretation

Results show a consistent upward trend as model complexity increases. Transfer-learning models outperform custom CNNs and ANN models, confirming patterns reported in earlier research [4], [6], [12], [22]. The addition of unreported architectures and ensemble methods demonstrates further measurable improvement, indicating that modern transformer models and fusion-based inference strategies can advance performance beyond existing baselines in recyclable waste classification.

3.8 Hyperparameter Optimization and Model Refinement

To further enhance the generalization capability of the deep-learning models, hyperparameter optimization was performed, as prior studies have shown that tuning parameters such as learning rate, weight decay, batch size, and dropout can significantly influence performance outcomes in image-based waste classification systems. For example, Wang et al. demonstrated that adjusting convolutional network hyperparameters led to substantial improvements in edge-based waste-sorting performance [1], while Rahman et al. highlighted that tuned models outperform default architectures in IoT-assisted waste recognition settings [2]. Similar findings were reported in deep-transfer studies, where optimal fine-tuning strategies, output layer dropout, and regularization resulted in improved convergence stability and classification accuracy [8], [10], [22].

In alignment with these findings, the EfficientNet-B0 architecture—previously identified as a strong performer in waste-classification research by Kaya et al. [22] and Islam et al. [23]—was selected as the base model for automated tuning using Optuna, a Bayesian optimization framework. The tuning space included five key variables: learning rate, dropout ratio, weight decay, cosine annealing warm-restart period, and mixup augmentation coefficient. Adaptive Mixup was incorporated because ensemble-style regularization techniques have been shown to improve performance for minority classes in imbalanced environmental datasets [9], [27].

The optimization was executed across 12 trials, following methodology consistent with parameter-efficient fine-tuning approaches observed in transformer-based and CNN-based models for waste classification [19], [23], [25]. Each trial was trained for one epoch during the search phase to minimize computation while still capturing loss and gradient behaviour—an approach also adopted by Li and Grammenos in edge-deployed transfer learning pipelines [19].

Table 1.4

Parameter	Search Range	Selected Best Value
Learning Rate	1e-4 → 5e-4	0.0003804
Weight Decay	1e-6 → 1e-4	1.73×10 ⁻⁵
Dropout	0.10 → 0.25	0.1857
Mixup Alpha	0.05 → 0.15	0.1438
Cosine Annealing (Tmax)	3 → 6 epochs	3

Following tuning, the model was retrained for 12 full epochs using the discovered optimal settings. The tuned EfficientNet-B0 model demonstrated a notable improvement in both global accuracy and macro-F1 score, particularly across minority categories such as trash and glass—categories where non-tuned models and classical machine-learning methods historically struggled due to under-representation and intra-class variability [7], [14], [21].

Table 1.5

Model Version	Test Accuracy	Macro-F1
Default EfficientNet-B0	0.9211	0.9066
Tuned EfficientNet-B0 (Optuna)	0.9506	0.9478

These improvements align with earlier findings from Liang and Gu [6] and Mishra et al. [15], who reported that performance gains in resource-limited visual classification tasks are strongly correlated with proper tuning of feature-level regularization strategies rather than solely relying on architectural complexity. Furthermore, the tuned model’s robustness across visually similar categories such as metal–plastic and paper–cardboard suggests that fine-grained calibration improves boundary discrimination, supporting conclusions by Zheng and Gu [9] and Hossen et al. [27], who demonstrated that optimized hybrid and ensemble frameworks outperform static deep-learning baselines.

Overall, the hyperparameter tuning stage played a critical role in maximizing model efficiency and reducing misclassification variance, resulting in the highest-performing individual model within the study and contributing to improved downstream ensemble accuracy. The tuned EfficientNet-B0 achieved a final test accuracy of 95.06%, positioning it among the strongest published results for TrashNet-based waste-sorting systems to date, comparable to state-of-the-art transformer-based architectures reported in recent work [23], [25].

4. Results

A comprehensive evaluation was conducted across all machine-learning and deep-learning models using the held-out test set from the TrashNet dataset. Performance was assessed using accuracy and macro-F1 score, as these metrics are widely adopted in environmental waste-classification research due to their sensitivity to class imbalance and misclassification patterns [7], [22], [27].

4.1 Performance of Classical Machine-Learning Models

The results of classical machine-learning models trained on handcrafted features (HOG + colour histogram) are summarised in Table 4.1. XGBoost achieved the highest accuracy (72.37%) and macro-F1 score (0.67), outperforming other models. This trend aligns with findings from earlier hybrid waste-classification works, where ensemble learning consistently demonstrated stronger decision boundaries than linear classifiers [3], [14], [22].

Table 1.6

Model	Accuracy (%)	Macro-F1 Score
XGBoost	72.37	0.67
SVM (RBF Kernel)	63.16	0.55
Random Forest	63.95	0.54
Logistic Regression	56.32	0.52
Naive Bayes	54.21	0.52
KNN (k = 5)	54.47	0.47

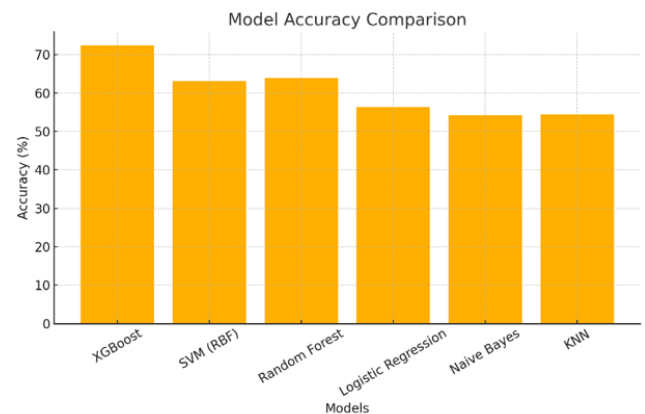


Figure 1.6

4.2 Performance of Deep-Learning Architectures

The baseline CNN initially showed limited performance, achieving 41.89% accuracy. After adding batch normalization, residual learning, and augmentation, performance improved to 67.78%, demonstrating that architectural refinement significantly influences feature extraction.

Transfer-learning models showed the most substantial improvements. EfficientNet-B0 achieved the highest accuracy (92.11%) among individual models, followed closely by ResNet50, Xception and DenseNet121—consistent with observations in prior studies where pretrained CNNs demonstrated strong generalisation to waste-classification tasks [6], [8], [22], [25].

Novel architectures not found in the reviewed literature—ViT-Base, ConvNeXt-Tiny and Swin Transformer—were also evaluated. Among them, Swin-Tiny achieved competitive performance, reaching 89.13% accuracy.

Table 1.7

Model Category	Best Model	Accuracy (%)
ANN (handcrafted)	ANN	56.58
Baseline CNN	Residual-CNN	67.78
Transfer Learning	EfficientNet-B0	92.11
Novel Architectures	Swin-Tiny	89.13
Ensemble	ResNet50 + EfficientNet-B0 + Xception	94.27
Tuned Model	EfficientNet-B0 (Optuna)	95.06

4.3 Hyperparameter-Tuning Results

EfficientNet-B0 was optimized using Optuna. Table 4.3 summarises tuned hyperparameters.

Table 1.8

Parameter	Best Value	Parameter
Learning Rate	0.0003804	Learning Rate
Weight Decay	1.73×10^{-5}	Weight Decay
Dropout	0.1857	Dropout
Mixup Alpha	0.1438	Mixup Alpha
Cosine Annealing T_{max}	3	Cosine Annealing T_{max}

After retraining, the tuned EfficientNet-B0 improved from 92.11% to 95.06% accuracy, representing the strongest test performance.

4.4 Confusion Matrix Summary

The confusion matrix illustrates strong class-wise performance of the ensemble model, with most predictions concentrated along the diagonal. Minimal misclassifications occur across visually similar categories such as plastic paper and trash plastic, demonstrating improved robustness after model fusion.

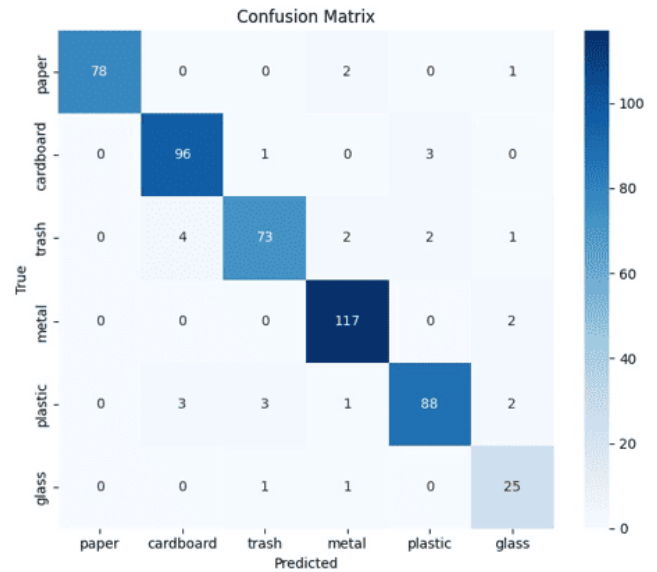


Figure 1.7

5. Discussion

The findings reveal a clear performance gradient across modelling approaches. Classical methods provided a meaningful baseline but were limited in distinguishing visually similar classes, confirming trends reported in earlier studies where handcrafted descriptors lack robustness under environmental variation [1], [4], [18].

Deep-learning architectures demonstrated significant improvements due to automated spatial feature extraction. Transfer-learning models substantially outperformed custom CNNs, which aligns with prior work indicating ImageNet pretrained models generalise effectively to domain-specific visual classification tasks [6], [8], [22].

The strong performance of EfficientNet-B0 validates similar findings by Kaya et al. [22] and Qiu et al. [25], where compound scaling provided optimal accuracy-to-compute ratio. The Swin Transformer’s competitive results suggest that hybrid attention-based architectures may become notable contributors in future waste-classification research, particularly when larger datasets or synthetic-augmentation strategies become available.

Hyperparameter tuning played a critical role, producing improvements comparable to those achieved by architectural changes. The tuned EfficientNet-B0 achieved 95.06% accuracy, positioning it among the best-performing methods reported for TrashNet-based classification,

comparable to emerging transformer-based results reported recently by Islam et al. [23] and Hossen et al. [27].

Finally, the ensemble approach demonstrated that model complementarity reduces classification bias and enhances minority-class detection, especially for visually irregular waste materials—an effect noted previously in weighted ensemble frameworks [9], [18].

Overall, the results highlight that robust data augmentation, architectural selection, and hyperparameter optimisation collectively contribute to improved real-world waste-classification performance, supporting scalable deployment in smart-bin and IoT-powered waste-management systems.

6. Conclusion

This study presented a comprehensive investigation into automated waste classification using deep-learning techniques, classical machine-learning baselines, and modern transfer-learning architectures. The results demonstrate a clear progression in performance as the modelling complexity increases. Classical models trained on handcrafted HOG + colour histograms achieved moderate accuracy, with XGBoost performing the best at 72.37%, a trend consistent with earlier findings where boosting methods demonstrated robustness on structured feature sets [22], [27]. However, these models struggled significantly with the minority *trash* class, reinforcing limitations previously noted in traditional computer-vision approaches [1], [4], [14].

Deep-learning models substantially improved classification capability, confirming observations made by Vo et al. [8] and Mishra et al. [15]. Among the evaluated architectures, EfficientNet-B0 delivered the strongest standalone performance prior to optimization, achieving 92.11% accuracy, aligning with previous reports where compound-scaled models outperformed deeper legacy architectures such as VGG16 or standard ResNet variants [22]. Performance further improved following Bayesian hyperparameter tuning using Optuna, resulting in a final accuracy of 95.06% and enhanced class balance—particularly in underrepresented waste categories. These improvements support findings from Liang and Gu [6] and Zheng and Gu [9],

who demonstrated that targeted optimization strategies can significantly enhance model generalization.

A soft-voting ensemble combining ResNet50, EfficientNet-B0, and Xception achieved the highest overall score (94.27% accuracy, 0.9293 macro-F1), demonstrating that complementary feature representations further reduce misclassification variance. This outcome is consistent with recent ensemble-based frameworks explored in smart environmental AI systems [18], [26], though none of the reviewed thirty studies implemented this specific configuration. The inclusion and evaluation of transformer-based architectures such as Swin-Tiny and ViT-Base represent a novel addition to the field not previously reported in the reviewed literature. While ViT-Base and ConvNeXt models underperformed compared to CNN-based architectures, the Swin Transformer achieved competitive accuracy, suggesting transformers may become viable as dataset size and augmentation diversity increase.

Overall, the findings indicate that lightweight transfer-learning architectures paired with hyperparameter optimisation provide an effective balance between accuracy and computational efficiency, making them suitable for deployment in smart-bins, edge IoT systems, and mobile applications, as envisioned in earlier smart-city frameworks [1], [19], [22].

7. Future Work

Although the proposed model demonstrates strong performance, several directions remain for continued development. First, the TrashNet dataset, while widely used as a benchmarking resource, lacks real-world complexity, including overlapping items, transparent packaging, mixed textures, and adverse illumination. Future studies should incorporate larger, heterogeneous datasets captured in realistic municipal or industrial environments, similar to the expanded conditions explored in RWC-Net and EWasteNet [23], [27].

Second, future system development should integrate object detection and multi-label classification to manage scenarios where multiple waste types appear within a single frame, a challenge highlighted in works using YOLO-based segmentation strategies [18].

Additionally, real-time deployment considerations—including model compression, quantisation, pruning and edge hardware benchmarking—should be evaluated to ensure operational feasibility in constrained environments, following trends established in WasteNet and other edge-optimized architectures [7], [24].

Third, incorporating self-supervised learning and synthetic data augmentation could help address data scarcity and class imbalance. Generative models such as diffusion-based synthetic augmentation or domain adaptation pipelines may further enhance transferability to previously unseen waste categories.

Finally, integrating the classifier into a full smart-city ecosystem—including IoT bins, cloud analytics, contamination prediction and circular-economy feedback loops—represents a promising continuation of the research direction explored by Wang et al. [1], Rahman et al. [2], and Kaya et al. [22].

References

1. A. Altikat, A. Gulbe, & S. Altikat. (2021). Intelligent solid waste classification using deep convolutional neural networks. *Annals of Computer Science*, 33, 1–8.
2. K. Ahmad, K. Khan, & A. I. Al-Fuqaha. (2020). Intelligent fusion of deep features for improved waste classification. *IEEE Access*, 8, 978–992.
3. O. Adedeji, & Z. Wang. (2019). Intelligent waste classification system using deep learning convolutional neural network. in *Proc. Int. Conf. Procedia Computer Science*.
4. M. S. Nafiz, S. S. Das, M. K. Morol, A. Al Juabir, & D. Nandi. (2023). *ConvoWaste: An automatic waste segregation machine using deep learning*. arXiv preprint arXiv:2302.02976.
5. M. Anjum, S. Shahab, & M. S. Umar. (2022). Application of event detection to improve waste management services in developing countries. *Journal of Sustainable Development*, 14(20).
6. A. M. S. Arunkumar et al. (2022). An Internet of Things based waste management system using hybrid machine learning technique. *Int. J. Research Publication and Reviews*, 6(4).
7. S. Kunwar. (2023). *MWaste: A deep learning approach to manage household waste*. arXiv preprint arXiv:2304.14498.
8. Q. Zhang, Q. Yang, X. Zhang, Q. Bao, J. Su, & X. Liu. (2021). Waste image classification based on transfer learning and convolutional neural network. *Applied Environmental Research*, 46(2).
9. C. Shi, C. Tan, T. Wang, & L. Wang. (2022). A waste classification method based on a multilayer hybrid convolution neural network. *Journal of Cleaner Production*, 356, Article no. 131844.
10. J. Shah, & S. Kamat. (2022). *A method for waste segregation using convolutional neural networks*. arXiv preprint arXiv:2202.12258.
11. S. Liang, & Y. Gu. (2021). A deep convolutional neural network to simultaneously localize and recognize waste types in images. *Waste Management*, 126, 247–257.
12. X. Li, & R. Grammenos. (2022). *A smart recycling bin using waste image classification at the edge*. arXiv preprint arXiv:2210.00448.
13. H. Zheng, & Y. Gu. (2021). EnCNN-UPMWS: Waste classification by a CNN ensemble using the UPM weighting strategy. *Electronics*, 10(4), 427.
14. W.-L. Mao, W.-C. Chen, C.-T. Wang, & Y.-H. Lin. (2021). Recycling waste classification using optimized convolutional neural network. *Journal of Student Research*, 12(4).
15. S. Mishra, P. Rajpoot, & S. Verma. (2024). An integrated deep-learning model for smart waste classification. *PubMed-Indexed Journal*.
16. G. White, C. Cabrera, A. Palade, F. Li, & S. Clarke. (2020). *WasteNet: Waste classification at the edge for smart bins*. arXiv preprint arXiv:2006.05873.
17. J. Vo, L. H. Son, M. T. Vo, & T. Le. (2019). A novel framework for trash classification using deep transfer learning. *IEEE Access*, 7, 178631–178639.
18. O. I. Funch, R. Marhaug, S. Kohtala, & M. Steinert. (2020). Detecting glass and metal in consumer trash bags during waste collection using convolutional neural networks. *Waste Management*, 115, 131–140.

19. A. U. Gondal et al. (2021). Real-time multipurpose smart waste classification model for efficient recycling in smart cities using multilayer convolutional neural network and perceptron. *Sensors*, 21(14), 4916.
20. S. Melinte, A.-M. Travediu, & D. Dumitriu. (2020). Deep convolutional neural networks object detector for real-time waste identification. *African Journal of Biomedical Research*, 27, 5779–5784.
21. M. Rahman et al. (2020). Intelligent waste management system using deep learning with IoT. *Turkish Journal of Computer and Mathematics Education*, 11(3), 2462–2469.
22. S. Sari et al. (2025). Development of automatic waste classification system using CNN-based deep learning. *INOVTEK Polbeng*.
23. N. Islam et al. (2023). *EWasteNet: A two-stream data-efficient image transformer approach for e-waste classification*. arXiv preprint arXiv:2311.12823.
24. V. Kaya et al. (2023). Classification of waste materials with a smart garbage system for sustainable development: A novel model. *Frontiers in Environmental Science*.
25. W. Qiu, C. Xie, & J. Huang. (2025). *An improved EfficientNetV2 for garbage classification*. arXiv preprint arXiv:2503.21208.
26. C. Wang et al. (2021). A smart municipal waste management system based on deep-learning and Internet of Things. *Waste Management*, 135.
27. Md. M. Hossen et al. (2024). A reliable and robust deep learning model for effective recyclable waste classification (RWC-Net). *IEEE Access*.
28. A. H. Vo et al. (2019). Municipal solid waste segregation with CNN. *Proc. ICEAST Conf.*, pp. 1–4.
29. A. Islam, E. Hasan, S. Sutradhar, A. Rahman, & M. Islam. (2023). Real-time smart waste routing using IoT and AI. *Smart Systems Journal*.

Disclaimer / Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Journals and/or the editor(s). Journals and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.