

---

Article

# Deep Ensemble Learning Model for Waste Classification Systems

Ahmet Alkilinc <sup>1,\*</sup>, Feyza Yıldırım Okay <sup>2</sup>, İbrahim Kök <sup>3</sup> and Suat Özdemir <sup>1</sup>

<sup>1</sup> Department of Computer Engineering, Hacettepe University, 06800 Ankara, Türkiye; ozdemir@cs.hacettepe.edu.tr

<sup>2</sup> Department of Computer Engineering, Gazi University, 06570 Ankara, Türkiye; feyzaokay@gazi.edu.tr

<sup>3</sup> Department of Artificial Intelligence and Data Engineering, Ankara University, 06100 Ankara, Türkiye; ikok@ankara.edu.tr

\* Correspondence: ahmetalkilinc@cs.hacettepe.edu.tr

## Abstract

Waste classification is a critical aspect of sustainable waste management systems. Traditional methods for waste classification are often inadequate to handle the complexity and diversity of materials encountered in real-world scenarios. This paper proposes novel deep ensemble learning models that combine pre-trained models with ensemble methods to improve waste classification performance. The proposed model leverages transfer and ensemble learning techniques, employing both averaging and weighted averaging methods to enhance waste classification accuracy. The proposed model is evaluated comprehensively on four publicly available waste image datasets containing various waste categories: TrashNet, TrashBox, Waste Pictures and Garbage Classification. The obtained results show that the averaging and weighted averaging ensemble methods improved classification accuracy by 1% to 3% over the strongest individual models. The weighted ensemble method achieves 96% accuracy, 94% precision, 97% recall and 95% F1 score on the TrashNet dataset. Statistical significance is verified using 5-fold cross-validation and paired *t*-tests ( $p < 0.05$ ). To ensure model explainability, the localization of important object regions is demonstrated with Grad-CAM visualizations. Overall, this study validates the potential of integrating deep image classification models with ensemble methodologies to improve the accuracy and efficiency of waste classification. The main contributions of this study can be summarized as follows: we design an efficient deep ensemble method that leverages multiple pre-trained models and ensemble techniques; we employ averaging and weighted averaging techniques to improve classification accuracy and model robustness; and lastly, we evaluate the model using multiple datasets to demonstrate generalizability, scalability and robustness.



Academic Editor: César Porras-Amores

Received: 21 November 2025

Revised: 13 December 2025

Accepted: 16 December 2025

Published: 19 December 2025

**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](#).

**Keywords:** deep learning; ensemble learning; transfer learning; waste classification; solid waste management; recyclable materials

---

## 1. Introduction

Rapid population growth and urbanization in cities are leading to an unprecedented increase in waste generation, posing significant challenges that necessitate a paradigm shift in conventional methods in resource management, environmental sustainability, and community resilience. According to recent studies, a 70% rise is expected in the global volume of waste by 2050 [1,2].

In this context, waste classification systems emerge as a critical subject, playing a significant role in mitigating environmental pollution, preserving resources, and fostering

efforts towards the waste recycling process. Traditional methods such as physical separation, chemical analysis, and visual inspection that rely on predefined rules and heuristic approaches are commonly used in waste classification [3]. However, the utilization of these methods can be challenging due to reasons such as the diversity of waste, human error, cost, time and identification of hazardous waste [4]. For these reasons, the development of automated waste classification systems integrated with modern technologies is crucial in overcoming these challenges and enhancing the efficiency of waste management processes. Recently, digital technologies such as the Internet of Things (IoT), Artificial Intelligence (AI) and sensor networks have started to play a major role in improving waste collection, sorting and management operations [5]. In addition, these technologies are widely used to improve operational efficiency, reduce collection route optimization and provide real-time decision-making.

The recent advances in Deep Learning (DL) and computer vision have made significant progress in the field of image classification. Many different methods are presented to enhance the accuracy and effectiveness of image classification. AlexNet [6], GoogLeNet [7] and ResNet [8] models have obtained remarkable achievements in large-scale image classification. Ioffe et al. [9] proposed Batch Normalization (BN) which has successfully alleviated the problem of gradient vanishing. Xie et al. [10] introduced ResNeXt that both reduced the number of parameters and further improved overall performance. DenseNet [11] addressed the problem of model performance degradation by implementing jump links. This resulted in significant improvements in both accuracy and speed compared to previous network architectures. EfficientNet [12] developed by Google achieved significant performance improvements.

DL has demonstrated outstanding achievements in various challenging computer vision domains and it has proven as a powerful paradigm for extracting robust and generalizable feature representations. Especially in the field of medical image analysis, DL architectures have performed successfully in critical tasks like disease detection and diagnosis [13–16]. In addition, DL methods have made significant advances in autonomous navigation systems in which robust feature extraction and decision-making processes are critical under dynamic and uncertain conditions [17,18]. These studies demonstrate the ability of DL models to process complex and highly variable data and provide a strong foundation for the application of DL models in waste classification where varied material types, inconsistent imaging conditions, and complex visual patterns are significant issues.

The advancements in DL techniques in recent decades have encouraged extensive research into various algorithms for waste classification. Sakr et al. [19] proposed an algorithm utilizing the AlexNet framework for waste image classification. A study by Yang et al. [20] conducted a comparison between Convolutional Neural Network (CNN)-based models and Support Vector Machines (SVMs) using the TrashNet dataset. Bircanoglu et al. [21] proposed the RecycleNet algorithm which is created by modifying the connection models of skip connections inside dense blocks. Kang et al. [2] presented a ResNet-34 based algorithm, incorporating three specific modifications: multiple feature fusion, reuse of features in the residual unit, and optimization of the activation function. Zeng et al. [22] proposed PublicGarbageNet, which is a multi-task classification algorithm where one task identifies four main categories of household waste and another task recognizes ten sub-classes of waste. Yang et al. [4] introduced GarbageNet, a new incremental learning framework for waste classification. Chen et al. [23] designed GCNet, which is a lightweight and effective waste classification model based on ShuffleNet v2.

Although DL significantly improves the waste classification, our literature review shows that most existing waste classification studies focus either on traditional Machine Learning (ML) models or approaches that utilize a single or individual pre-trained DL

model. However, these models suffer from low accuracy and poor generalization issues when trained on small-sized and imbalanced datasets. The main objective of this study is to address these issues in waste classification systems; while many pre-trained CNN models are applied separately for waste classification, there is a significant gap in evaluating how ensemble learning strategies can effectively leverage the complementary strengths of multiple models to improve the performance of waste classification. Therefore, we take advantage of average and weighted average ensemble methods by combining high-performance pre-trained CNN models. Ensemble methods [24] have the potential to improve the accuracy of DL models. In the context of waste classification, employing an ensemble method can effectively boost classification accuracy by leveraging the strengths of multiple models. This collaborative approach enables the development of a more robust model by utilizing each model's individual strengths while mitigating their weaknesses. Accordingly, first, the waste classification models are trained using transfer learning with various pre-trained models such as DenseNet121, DenseNet201, MobileNetV2, MobileNetV3Large, InceptionV3, ResNet50, ResNet50V2, ResNet101, ResNet101V2, Xception, ConvNeXtTiny, ConvNeXtLarge, EfficientNetB0, EfficientNetB7, EfficientNetV2B0, EfficientNetV2L. Then, the three models are selected from these trained models to implement averaging and weighted averaging ensemble learning methods. In this study, we propose the combination of pre-trained deep models with ensemble methods to improve accuracy while preserving robustness. We provide a comprehensive benchmark analysis on four publicly available waste datasets and present an optimized weighted ensemble strategy which improves the prediction performance while maintaining practical applicability. The main contributions of the paper are as follows:

- We present efficient deep ensemble learning models, integrating pre-trained models with ensemble methods to provide more accurate results for waste management systems.
- We perform a comparative evaluation of the waste classification problem using sixteen different pre-trained DL models on four waste datasets.
- We provide a detailed overview of the existing studies on waste classification.
- We implement Grad-CAM method to ensure the explainability of models within the waste classification task.

This paper is structured as follows. Section 2 presents a review of the literature on waste classification. Section 3 describes the pre-trained models, ensemble learning methods and also describes the proposed waste classification method. Section 4 presents the dataset, performance metrics and experimental study and results. Finally, Section 5 provides general concluding remarks.

## 2. Related Work

Enhancing waste management practices can facilitate waste recycling and ensure efficient and economical use of resources. Research in this area covers a wide range of studies investigating various methodologies. Traditional approaches are largely based on manual sorting techniques which are labor-intensive and error-prone. However, recent developments in computer vision, ML and DL have significantly contributed to the waste classification process. These developments have resulted in extensive research targeting waste classification. The related studies in waste classification field are summarized and grouped by their dataset and methodology in Table 1.

Yang et al.'s [20] hand-collected, publicly available dataset with approximately 400–500 images per class is created to classify waste into recycling category by providing an effective waste processing method. Utilizing SVM with scale-invariant feature transformation (SIFT) features and CNN, their study compares their performance. Exper-

imental results indicate that SVM reaches 63% accuracy. Zhihong et al. [25] introduced a robotic grasping system designed for automatic waste sorting using ML methods. To achieve precise grasping, they used VGG-16 model and Region Proposal Generation (RPN) for pose estimation and object recognition. Bai et al. [26] introduced an intelligent waste collection robot capable of detecting and operating autonomously. The robot utilizes a deep neural network for both waste recognition and ground segmentation. They achieved 95% accuracy in waste recognition using ResNet-34. Chu et al. [27] suggested a multi-layer hybrid DL system (MHS) that combines high-resolution cameras and sensors to classify urban waste automatically. With this approach, they achieved over 90% accuracy in distinguishing recyclable items from non-recyclable items and outperformed the traditional CNN method based on images alone. Rabano et al. [28] introduced a model for waste classification using the MobileNet model on the Trashnet dataset. They utilized transfer learning from a pre-trained model on ImageNet. The suggested model achieved a test accuracy of 87.2%.

Satvilkar et al. [29] addressed the important topic of waste classification, highlighting its importance in waste management and the unique role of image processing through analytics. The paper explores multiple methods to determine the most efficient approach or combination for image classification using methods such as CNN, SVM, XGB, RF, KNN. CNN achieved the best classification accuracy of 89.81%. Bircanoglu et al. [21] proposed RecycleNet which is a deep CNN architecture optimized to classify recyclable objects by modifying skip connections within dense blocks. The proposed model decreased the number of parameters in a 121-layer network to about 3 million and achieved 81% accuracy on the TrashNet dataset. Zeng et al. [30] presented a multiscale convolutional neural network (MSCNN) method that uses aerial hyperspectral data for large-area waste detection. The proposed algorithm shows strong performance in litter detection and outperforms existing hyperspectral image (HSI) classification methods on publicly available datasets such as Indian Pines and Pavia University. Adeiji et al. [31] suggested a method for waste classification utilizing the ResNet-50 CNN model as a feature extractor and SVM for classification. This approach obtained an accuracy of 87% on the TrashNet dataset. Ruiz et al. [32] trained and evaluated different DL architectures with the TrashNet dataset to classify waste. Specifically, they compared CNN models including ResNet, Inception and VGG. The highest accuracy is achieved from the combined Inception-ResNet model, which achieves 88.6% accuracy.

Endah et al. [33] suggested a method for waste classification using transfer learning on the TrashNet dataset. They applied Xception, VGG16 and ResNet-50 models. The results demonstrate that the Xception model attains the best accuracy of 88%. Toğacar et al. [34] proposed a model that reconstructs the used dataset with AutoEncoder network and extracts feature sets using CNN architectures. In the proposed model, these extracted features are subsequently merged and Ridge Regression (RR) is applied to decrease the number of features. SVM is used as the classifier and the classification accuracy is 99.95% in experiments on a two-class open-access dataset. Azis et al. [35] presented a DL and computer vision-based method for waste classification using Inception-v3. The proposed model achieved 92.5% accuracy on TrashNet. Shi et al. [36] addressed challenges in waste classification by proposing a technique which is using a multilayer hybrid convolution neural network (MLH-CNN). This method, resembling VggNet yet with a simpler design and fewer parameters, aims to enhance accuracy and reduce computational time. Experimental findings on the TrashNet dataset demonstrate a classification accuracy of up to 92.6%. Zhang et al. [37] proposed a DL-based classification model for recyclable waste images, which integrates a self-monitoring module into the residual network model to improve feature representation and extract features from various waste images. The proposed model

achieved an accuracy of 95.87%. Majchrowska et al. [38] introduced new datasets, classify-waste and detect-waste, which combine open-source datasets with merged annotations covering various waste categories. They also introduced a two-stage detector that uses EfficientDet-D2 for waste detection and EfficientNet-B2 for its classification. Kumsetty et al. [39] introduced TrashBox, which is a new waste dataset consisting of seven classes. In addition, a new DL framework utilizing quantum transfer learning is explored and promising results are obtained in experimental evaluations on different datasets. Zhou et al. [40] presented a DL method for the classification of medical waste. The suggested classification method uses ResNeXt as a deep neural network and achieves remarkable results using transfer learning methods. In the experiments, it correctly classifies eight types of medical waste in the medical dataset with an accuracy of 97.2%.

**Table 1.** Summary of the literature on waste classification.

Ref	Year	Task	ML/DL Model(s)	Dataset	#Class	Performance Measure
[20]	2016	Waste classification	SIFT + SVM	TrashNet	6	Accuracy
[25]	2017	Waste sorting	VGG-16	Own data	1	Miss rate False rate
[26]	2018	Waste recognition	ResNet-34	Own data	6	Accuracy
[27]	2018	Waste classification	CNN, MLP, AlexNet	Own data	2	Accuracy, Precision, Recall
[28]	2018	Waste classification	MobileNet	TrashNet	6	Accuracy
[29]	2018	Waste classification	CNN, SVM, XGB, RF, KNN	TrashNet	6	Accuracy, Precision, Recall, F1-score
[21]	2018	Waste classification	ResNet50	TrashNet	6	Accuracy
[30]	2019	Waste detection, HSI classification	Multi-Scale CNN,	Indian Pines dataset, Pavia University dataset	16	Overall accuracy, Average accuracy, Kappa coefficient
[31]	2019	Waste classification	ResNet-50, SVM	TrashNet	6	Accuracy
[41]	2019	Waste classification	VGG, Inception, ResNet	TrashNet	6	Accuracy
[34]	2020	Waste classification	AutoEncoder, AlexNet, GoogLeNet, ResNet-50, SVM	Waste Classification data	2	Accuracy, Precision, F1-score
[33]	2020	Waste classification	VGG16, ResNet-50, Xception	TrashNet	6	Accuracy, Precision, Recall
[35]	2020	Waste classification	Inception-v3	TrashNet	6	Accuracy
[42]	2021	Waste classification	InceptionV3	Garbage Classification	12	Accuracy, Precision, Recall, F0.5-score
[36]	2021	Waste classification	MLH-CNN, VGG16, AlexNet, ResNet50	TrashNet	6	Accuracy, Precision, Recall, F1-score
[37]	2021	Waste classification	ResNet18	TrashNet	6	Accuracy, Precision, Recall, F1-score
[38]	2022	Waste detection Waste classification	EfficientDet-D2, EfficientNet-B2	Detect-waste Cassify-waste	8	mAP, Precision, Recall, F1-score
[40]	2022	Waste classification	ResNeXt-50	Medical waste dataset	8	Precision, Recall, F1-score
[39]	2022	Waste detection Waste classification	ResNet-34, ResNet-50, ResNet-101, VGG-19, DenseNet-121	TrashNet TACO TrashBox	6 28 7	Accuracy
[43]	2023	Waste classification	GoogleNet, ResNet, DenseNet, ResNeXt, EfficientNet	Garbage Classification	12	Accuracy, Precision, Recall, F1-score
[44]	2023	Waste classification	VGGNet16, Resnet50, MobileNetV2, InceptionV3,CNN	Garbage Classification	8	Accuracy, Precision, Recall, F1-score
[45]	2023	Waste classification	ResNet18	TrashNet	6	Accuracy, Precision, Recall, F1-score
[46]	2023	Waste detection Waste classification	MobileNet-v2	HUAWEI-40	4	Accuracy, Precision
[47]	2023	Waste classification	ResNet-34, ResNet-101, VGG-16	TrashNet, TACO	6	Accuracy
[48]	2024	Waste classification	GoogleNet, ResNet50, Inception-v3, MobileNet-v2, DenseNet201.	TrashNet	6	Accuracy, Precision, Recall, Specificity, F1-score
[49]	2024	Waste classification	VGG-16, ResNet-34, ResNet-50, AlexNet, LSTM	Recycle Waste image dataset	2	Accuracy, Precision, Recall, F1-score
[50]	2024	Waste classification	VGG19	TrashNet, GarClass	6	Accuracy, Precision, Recall, F1-score

**Table 1.** Cont.

Ref	Year	Task	ML/DL Model(s)	Dataset	#Class	Performance Measure
[51]	2024	Waste classification	YOLO 5, YOLO 7	e-waste dataset	5	mAP, Accuracy, Precision, Recall, F1-score
[52]	2024	Waste classification	VGG16	TrashNet, Own data	6	Accuracy
[53]	2025	Waste classification	CNN, YOLO	Own data	4	Accuracy, Precision, Recall, F1-score

In another study, Shukurov [43] proposed an algorithm using fine-tuned CNN (GoogleNet, DenseNet, ResNet, ResNeXt, EfficientNet). Its experiments are conducted on an open-source dataset of 12 categories using SGD and Adam optimization algorithms. The experimental results show that fine-tuned ResNeXt model reaches 95% accuracy. Chen et al. [42] presented an algorithm based on InceptionV3 networks for waste classification. They conducted their experiments on the open-source Garbage Classification (12 classes) dataset on Kaggle using transfer learning. The suggested model achieved 93.1% accuracy in waste classification. Dey et al. [44] presented a custom CNN model designed for waste classification. In the study, pre-trained models (ResNet50, VGGNet16, InceptionV3, MobileNetV2) are compared with a proposed custom CNN model. The suggested model achieved 97.58% accuracy. Huang et al. [45] proposed ResMsCapsule network, a capsule network-based waste classification model. This model improves the performance of the basic capsule network by combining the residual network and the multiscale module. Experiments with the TrashNet dataset showed that ResMsCapsule has a classification accuracy of 91.41%. Jin et al. [46] developed a new machine vision system based on DL and transfer learning for waste classification. The proposed model is based on MobileNetV2. The model uses an attention mechanism and transfer learning to improve the accuracy. Dimension reduction with PCA was applied to enable the model to run in real time on edge devices. In the experiments, 90.7% accuracy was achieved on the Huawei Cloud dataset. Kumsetty et al. [47] proposed a computer-vision-based automatic waste classification method. In the study, the quality of the datasets was improved by data augmentation and image processing techniques and transfer learning-based models such as ResNet and VGG were used. In the experiments on TrashNet and TACO datasets, 93.13% and 91.34% accuracy was achieved, respectively.

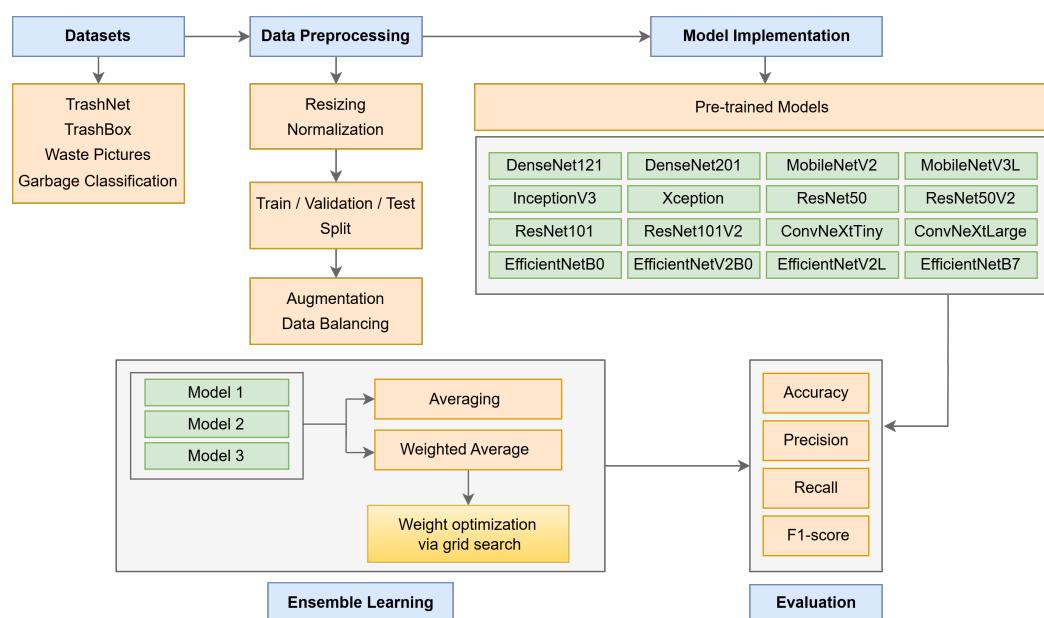
Hossen et al. [48] introduced RWC-Net, which is based on a combination of DenseNet201 and MobileNet-V2 for the classification of recyclable waste. This model uses the feature extraction and learning capabilities from both pre-trained models. The model accomplishes an overall accuracy of 95.01% on the TrashNet dataset. Lilhore et al. [49] proposed a waste classification model that combines CNN and long short-term memory (LSTM). The suggested model also uses a transfer learning method that includes the benefits offered by ImageNet. This model's performance was evaluated on a two-class dataset and the proposed model achieved a precision of 95.45%. Quan et al. [50] proposed a novel method for waste classification in IoT environments that prioritizes privacy preservation. The method combines differential privacy, federated learning and transfer learning techniques to enable collaborative model training while preserving sensitive data. Sarswat et al. [51] proposed a new method of electronic waste classification using DL and computer vision. Researchers have developed real-time object recognition systems for e-waste classification using YOLO 7 and 5 algorithms. At the end of experimental studies, the YOLO 7 model achieved 94% prediction accuracy. Lin et al. [52] presented a multiscale feature fusion strategy (MFFS) for the classification of real-world waste with similar shape, texture or contaminated appearance. This method improves classification accuracy by combining local fine-grained details and global coarse-grained features. The presented

model achieved 94.1% accuracy on the TrashNet dataset and 95.5% accuracy on the custom dataset. Kumar et al. [53] presented an innovative framework integrating AI and IoT technologies to enable the safe and efficient management of hazardous waste in hospitals. In this study, IoT-enabled smart waste bins ensure a secure waste collection process, while CNN and Adaptive Neuro Fuzzy Inference System (ANFIS) increase waste detection and classification accuracy. The preliminary experimental results obtained show a reduction in manual intervention and a significant increase in classification performance.

While existing studies in the literature have demonstrated the potential of DL models for automatic waste classification, there are critical gaps in the field that remain unaddressed. First, most studies are performed on a single dataset or limited data conditions which limits the generalizability of the proposed models to various waste environments. Second, the existing studies focus mainly on traditional ML models or a single pre-trained DL approach. However, these models demonstrate low accuracy and poor generalization performance when trained on small and imbalanced datasets. Although many pre-trained CNN models for waste classification are applied individually in the literature, there is a significant gap in evaluating how ensemble learning strategies can effectively leverage the complementary strengths of multiple models to improve waste classification performance. This study aims to address the aforementioned methodological and practical gaps in the waste classification by integrating optimized ensemble learning methods with experiments on multiple datasets.

### 3. Proposed Deep Ensemble Model for Waste Classification

This section overviews pre-trained models trained on large image datasets and explores ensemble learning methods. Furthermore, we introduce our proposed deep ensemble model architecture. The methodological workflow diagram is shown in Figure 1.



**Figure 1.** Workflow of the proposed methodology.

#### 3.1. Pre-Trained Models

CNNs are widely used for image-based feature extraction and spatiotemporal forecasting [54,55]. Traditional ML techniques require manually prepared feature extraction process and usually creating difficulties in generalizing on different image conditions such as lighting changes, texture variations and background complexity. On the other hand, CNNs automatically learn discriminative features directly from the raw images, enhancing

the classification robustness and scalability. Residual learning-based architectures such as ResNet help prevent vanishing gradients through skip connections and enable deeper networks to be trained effectively [8]. ConvNeXt [56] and EfficientNet [12] architectures further improve performance by utilizing modern design techniques such as compound scaling and improved convolutional blocks. However, CNN models still suffer from challenges caused by dataset variability, class imbalance and overfitting, particularly in waste classification where object contamination and deformation are common. To address these challenges, we introduce an ensemble-based approach by combining the predictions of multiple models with the goal of achieving better classification accuracy.

When selecting pre-trained models for our method, we prioritized architectures that have shown strong performance on large-scale benchmarks such as ImageNet and proven effectiveness in waste classification. The selected models represent a variety of architectural designs, including CNNs with different depths, residual connections, and initialization modules. This diversity enables the proposed ensemble model to capture a wide range of feature representations and this is crucial given the complex and heterogeneous nature of waste images which differ in texture, shape and background complexity. Utilizing such complementary models allows our model to better generalize across different waste types and environmental conditions.

InceptionV3 [57] is a CNN architecture that uses inception modules with various kernel sizes to capture features of different scales. It efficiently learns fine and complex patterns by factorizing convolutions into smaller operations.

ResNet-50 and ResNet-101 [8] are deep residual networks with 50 and 101 layers. Their skip connections help the model learn residual functions instead of full mappings, allowing deeper architectures to be trained while mitigating vanishing gradients. These models strike a balance between model complexity and computational efficiency, making them popular choices for various computer vision tasks, including waste classification.

DenseNet-121 [11] is a CNN architecture characterized by its dense connectivity, in which every layer is directly linked to each other layer in a feed-forward. This design produces compact and expressive models. There are different variants like DenseNet-121, DenseNet-169 and DenseNet-201, which offer different levels of depth and accuracy.

Xception [58] is a deep CNN architecture that emphasizes depthwise separable convolutions. It reduces computation by separating spatial and channel operations while capturing complex structures. Furthermore, it performs well on challenging tasks and is suitable for environments with limited processing power.

MobileNetV2 [59] is a lightweight architecture designed for mobile and edge devices. It incorporates separable convolutions in the depth direction and uses residuals and linear bottlenecks to achieve superior performance on resource-constrained devices. Its small size and low latency make it ideal for real-time, resource-constrained applications.

EfficientNet [12] is a CNN architecture that systematically scales the model's depth, width and resolution to optimize efficiency and accuracy. The models range from EfficientNetB0 to EfficientNetB7, allowing users to choose architectures based on their computational needs.

ConvNeXt [56] is based on ResNet-style architectures and incorporates modern design principles such as improved convolutional operations and scaling strategies. These enhancements make ConvNeXt models both accurate and computationally efficient, providing flexibility in model size, performance and resource usage. A summary of the pre-trained models is shown in Table 2.

**Table 2.** The features of pre-trained models.

Model	Year	Parameter	Depth	Complexity
InceptionV3	2015	23.9 M	Deep	Complex
ResNet50	2015	25.6 M	Deep	Moderate
ResNet50V2	2016	25.6 M	Deep	Moderate
ResNet101	2015	44.7 M	Deep	Complex
ResNet101V2	2016	44.7 M	Deep	Complex
DenseNet121	2016	8.1 M	Deep	Complex
DenseNet201	2017	20.2 M	Deep	Complex
Xception	2017	22.9 M	Deep	Complex
MobileNetV2	2018	3.5 M	Shallow	Lightweight
MobileNetV3L	2019	3.9 M	Shallow	Lightweight
EfficientNetB0	2019	5.3 M	Moderate	Efficient
EfficientNetB7	2019	66.7 M	Moderate	Efficient
EfficientNetV2B0	2021	7.2 M	Moderate	Efficient
EfficientNetV2L	2021	119.0 M	Moderate	Efficient
ConvNeXtTiny	2022	28.6 M	Moderate	Complex
ConvNeXtLarge	2022	197.7 M	Moderate	Complex

### 3.2. Ensemble Learning Methods

Ensemble learning [24] is a ML technique aimed at enhancing the accuracy and robustness of predictions by aggregating the outputs of multiple models. By leveraging the collective intelligence of these models, ensemble learning reduces the impact of errors or biases that might be present in individual models. This method improves both accuracy and robustness to uncertainties in the data. The effectiveness of ensemble learning in combining predictions from multiple models has made it a powerful tool across various fields. Ensemble learning is used in different areas, such as image classification [60,61], text classification [62,63], intrusion detection [64], and activity recognition [65].

The averaging ensemble method [24] simply averages the predictions of individual models to obtain the final prediction. If you have  $N$  models, each providing a prediction  $y_i$  for a given input, the ensemble prediction  $y_{\text{ensemble}}$  is calculated as the average:

$$y_{\text{ensemble}} = \frac{1}{N} \sum_{i=1}^N y_i \quad (1)$$

Averaging is commonly employed when all models are assumed to contribute equally to the final prediction.

The weighted averaging ensemble method [24] assigns different weights to the predictions of individual models before averaging them. The weights reflect the importance or confidence assigned to each model. If  $w_i$  represents the weight assigned to the prediction of model  $i$ , the weighted ensemble prediction  $y_{\text{ensemble}}$  is calculated as follows:

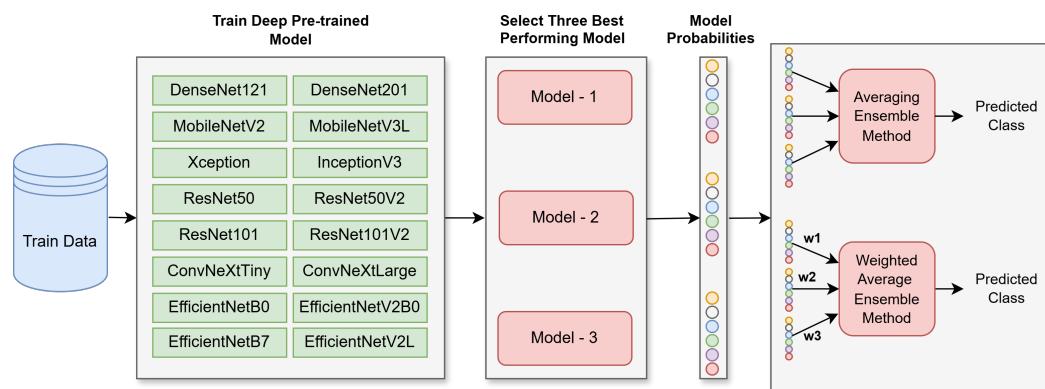
$$y_{\text{ensemble}} = \frac{\sum_{i=1}^N w_i \cdot y_i}{\sum_{i=1}^N w_i} \quad (2)$$

The weighted averaging ensemble method is beneficial when some models are considered more reliable or accurate than others. Assigning higher weights to more trustworthy models allows them to have a more significant influence on the final prediction.

### 3.3. Proposed Deep Ensemble Model

In this section, we elaborate on the experimental methodology employed in our study, which involves the utilization of pre-trained DL models and ensemble learning methods to enhance the overall model performance.

Firstly, we conduct experiments using sixteen different pre-trained DL models. These models are selected based on their built-in capabilities in various computer vision tasks. Each model is applied individually to the dataset to extract features and make predictions. Then, a comprehensive evaluation is performed to assess their respective performances. For the evaluation, we use performance measures including loss value, accuracy, precision, and recall. Then, we select three models and use ensemble learning techniques to leverage the effectiveness of the selected models and improve the overall performance. We apply two ensemble methods, averaging and weighted average, for combining the predictions from the selected models. The proposed models are shown in Figure 2.



**Figure 2.** Proposed model.

In the averaging ensemble method, the predictions of the three models are combined and averaged to obtain a prediction. This method assumes that each model contributing to the final prediction is of equal importance. In the weighted averaging ensemble method, it is calculated by assigning different weights to the predictions of each model. The weights are determined according to the performance and reliability of each model, thus more effective models have a greater effect on the prediction. In order to determine the optimal weights for the weighted averaging ensemble, we performed a grid search optimization over the possible combinations of weights assigned to each model. Let  $m_1, m_2, m_3$  be the three pre-trained models utilized in the ensemble. The final prediction  $P$  is calculated as follows:

$$P = \sum_{i=1}^3 w_i \cdot p_i \quad (3)$$

where  $p_i$  is the probability vector output of model  $m_i$  and  $w_i$  is the corresponding weight. In order to identify the best-performing weight configuration, a grid search was conducted by iterating  $w_1, w_2, w_3 \in \{1, 2, \dots, 9\}$ . To ensure that the weights form a valid probability distribution, they were normalized such that:

$$\tilde{w}_i = \frac{w_i}{w_1 + w_2 + w_3}, \quad \text{for } i \in \{1, 2, 3\}, \quad (4)$$

where  $\tilde{w}_i$  denotes the normalized weight assigned to model  $m_i$ . After evaluating all valid weight combinations, the weights that produce the highest classification accuracy on the test set are selected for the final test. This tuning is performed separately for each dataset to ensure the best performance. The pseudo-code of the proposed weighted averaging ensemble method is given in Algorithm 1.

**Algorithm 1** weighted averaging ensemble method

---

```

1: Generate predictions from selected three models:  $preds_1, preds_2, preds_3$  and true labels
    $labels$ 
2: Initialize  $max\_accuracy \leftarrow 0$ 
3: Initialize  $best\_weights \leftarrow [0, 0, 0]$ 
4: for  $w_1$  from 1 to 9 do
5:   for  $w_2$  from 1 to 9 do
6:     for  $w_3$  from 1 to 9 do
7:       Compute total weight:
8:        $total \leftarrow w_1 + w_2 + w_3$ 
9:       Normalize the weights:
10:       $w \leftarrow [w_1 / total, w_2 / total, w_3 / total]$ 
11:      Compute weighted predictions:
12:       $wted\_preds \leftarrow \sum_{i=1}^3 w_i \cdot preds_i$ 
13:      Determine final predictions:
14:       $final\_preds \leftarrow \text{argmax}(wted\_preds, axis = 1)$ 
15:      Compute accuracy:
16:       $accuracy \leftarrow \text{AccuracyScore}(labels, final\_preds)$ 
17:      if  $accuracy > max\_accuracy$  then
18:         $max\_accuracy \leftarrow accuracy$ 
19:         $best\_weights \leftarrow (w_1, w_2, w_3)$ 
20:      end if
21:    end for
22:  end for
23: end for
24: return  $best\_weights$ 

```

---

## 4. Experimental Results and Evaluation

### 4.1. Datasets

We evaluated the proposed method using four distinct datasets, namely TrashNet [66], TrashBox [39], Waste Pictures [41] and Garbage Classification [67]. The statistical information of the datasets is shown in Table 3. The data sets are divided into 80% training, 10% validation and 10% testing. In the pre-processing step, input images are resized to  $224 \times 224$  pixels to match the pre-trained models' input size. We apply the pre-process\_input function provided by the framework of each pre-trained model (e.g., Keras) for proper normalization and scaling of input data. We used several data augmentation techniques during training to make the model more generalizable and reduce the overfitting. These included random rotations by at most  $30^\circ$ , horizontal and vertical shifts in the range 0.1, zoom transformations in the range 0.1, horizontal flipping and brightness changes in the range [0.8, 1.2].

**Table 3.** Statistics of waste datasets.

Dataset	# Images	# Classes	Class Imbalance	Type	Annotation
TrashNet	2527	6	Low	Classification	Clear background
TrashBox	17,853	7	Low	Classification	Scraped from web
Garbage Classification	15,515	12	High	Classification	Scraped from web
Waste Pictures	23,087	34	High	Classification	Scraped from Google search

We analyzed the class distributions of all datasets to address class imbalance issues and used oversampling and class weighting methods to handle this problem. In the TrashNet and TrashBox datasets, there is no significant class imbalance, so we applied oversampling to increase the number of samples from the minority classes. On the other hand, the Garbage Classification and Waste Pictures datasets have higher class imbalance due to the significant differences in the number of categories and the number of images per class.

In order to mitigate this issue without excessively increasing the dataset size, we used class weighting during model training. To maintain the integrity of the evaluation and prevent any train-test contamination, all data augmentation and oversampling operations are applied after the dataset is split into training, validation and test subsets and only to the training data. Table 4 shows the number of training, test and validation samples before oversampling, while Table 5 shows the number of training, test and validation samples after oversampling.

**Table 4.** Number of images in the training, validation and test sets before oversampling.

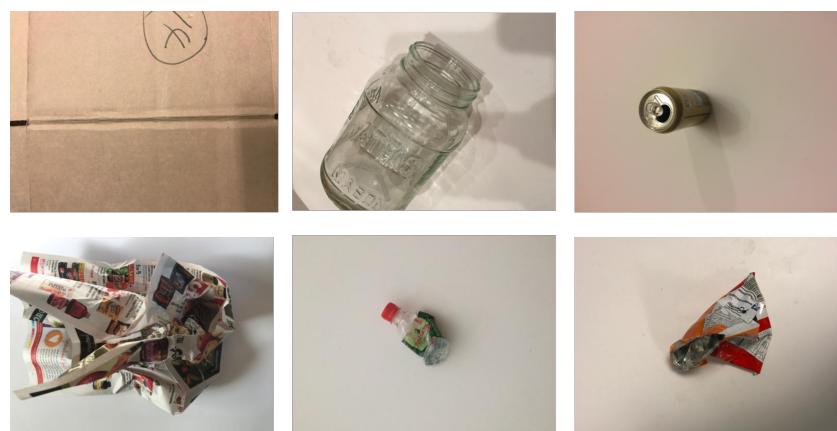
Dataset	Train	Validation	Test	Total
TrashNet	2021	253	253	2527
TrashBox	14,279	1781	1793	17,853
Waste Pictures	14,268	3567	5252	23,087
Garbage Classification	12,412	1551	1552	15,515

**Table 5.** Number of images in the training, validation and test sets after oversampling.

Dataset	Train	Validation	Test	Total
TrashNet	3564	253	253	4070
TrashBox	16,835	1781	1793	20,409
Waste Pictures	14,268	3567	5252	23,087
Garbage Classification	12,412	1551	1552	15,515

#### 4.1.1. TrashNet

The TrashNet [66] dataset consists of six common waste categories, namely cardboard, glass, plastic, paper, metal, and trash. The images were taken by positioning the object on a white poster board and utilizing natural sunlight and ambient room lighting. They were subsequently resized to  $512 \times 384$ . The dataset contains a total of 2527 waste images in jpg format. Example images from the dataset are shown in Figure 3.



**Figure 3.** Sample data images in the TrashNet dataset.

#### 4.1.2. TrashBox

The TrashBox [39] dataset is a comprehensive collection of images specifically organized for waste classification and detection research. It contains a total of 17,853 waste images in seven classes (plastic, glass, metal, cardboard, e-waste, paper, medical waste). It is a valuable resource for researchers interested in developing and evaluating ML and DL models for automatic waste detection and classification.

#### 4.1.3. Waste Pictures

The Waste Pictures [41] dataset consists of approximately 23,087 waste images from Google search results. The dataset contains 34 different image types, including organic, recyclable, and hazardous waste, which gives the dataset significant diversity.

#### 4.1.4. Garbage Classification

The Garbage Classification [67] dataset is an open-source dataset created on Kaggle using web scraping. The dataset contains 15,515 images from 12 different classes of household waste: cardboard, paper, biological, plastic, metal, brown glass, green glass, white glass, shoes, clothes, trash, and batteries.

### 4.2. Performance Metrics

The model performance is evaluated by measures like accuracy, precision, recall, and F1 score.

Accuracy refers to how close a measurement is to the true or accepted value. It is calculated by dividing the number of correctly predicted samples by the total number of samples.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}} \quad (5)$$

Precision quantifies the accuracy of the model's positive predictions. It is calculated as the ratio of correctly predicted positive observations to the total number of positive predictions.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (6)$$

Recall, also known as sensitivity or true positive rate, measures the capability of the model to capture all of the relevant positive samples. It is the ratio of accurately predicted positive observations to the total actual positives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (7)$$

F1-score is the harmonic average of precision and recall. It gives a balanced evaluation of a model's performance by taking into account both false positives and false negatives.

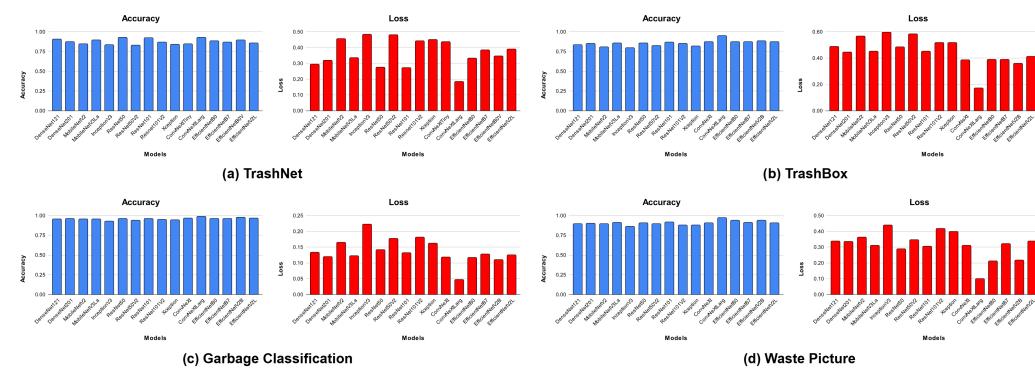
$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

### 4.3. Experimental Results

The experiments are conducted to compare the performance of various pre-trained models in waste classification and to develop a better model with ensemble learning methods. Table 6 summarizes the main findings of the pre-trained models. Furthermore, Figure 4 graphically shows the results. The experiments are performed on Google Colab using the Keras library with the TensorFlow backend. We performed hyperparameter tuning using the KerasTuner framework to optimally configure the base models. A grid search was performed to choose the best optimizer from {Adam, RMSprop, SGD} and to find the learning rate from {0.01, 0.001, 0.0001} for each model. The epoch number was set to 50, the batch size to 32, and early stopping (patience = 10) was applied to prevent overfitting. The selected optimizers and learning rates for each pre-trained model after hyperparameter tuning are shown in Table 7.

**Table 6.** Loss and accuracy of the pre-trained models.

Dataset		TrashNet		TrashBox		Waste Pictures		Garbage Classification	
Model		Loss	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy
DenseNet121		0.29	0.91	0.49	0.84	0.34	0.90	0.13	0.95
DenseNet201		0.32	0.87	0.44	0.85	0.33	0.90	0.12	0.96
MobileNetV2		0.45	0.85	0.57	0.81	0.36	0.89	0.17	0.95
MobileNetV3L		0.34	0.89	0.45	0.86	0.31	0.91	0.12	0.96
InceptionV3		0.48	0.83	0.60	0.80	0.44	0.87	0.22	0.93
ResNet50		0.27	0.93	0.49	0.86	0.29	0.91	0.14	0.96
ResNet50V2		0.48	0.83	0.58	0.82	0.35	0.90	0.18	0.94
ResNet101		0.27	0.92	0.45	0.87	0.30	0.92	0.13	0.96
ResNet101V2		0.44	0.87	0.52	0.85	0.42	0.88	0.18	0.95
Xception		0.45	0.84	0.52	0.82	0.40	0.88	0.16	0.94
ConvNeXtTiny		0.44	0.85	0.39	0.87	0.31	0.91	0.12	0.97
ConvNeXtLarge		0.18	0.93	0.17	0.95	0.10	0.97	0.05	0.99
EfficientNetB0		0.33	0.89	0.39	0.87	0.21	0.94	0.12	0.96
EfficientNetB7		0.38	0.87	0.39	0.87	0.32	0.91	0.13	0.96
EfficientNetV2B0		0.35	0.89	0.36	0.88	0.22	0.94	0.11	0.98
EfficientNetV2L		0.39	0.86	0.41	0.87	0.34	0.90	0.13	0.97

**Figure 4.** Loss and accuracy of the models on different datasets. (a) TrashNet, (b) TrashBox, (c) Garbage Classification, and (d) Waste Pictures dataset.**Table 7.** Summary of the CNN models that are trained on all datasets.

Model	Optimizer	LR	# Parameters
DenseNet121	RMSprop	$1 \times 10^{-3}$	7,044,679
DenseNet201	RMSprop	$1 \times 10^{-3}$	18,335,431
MobileNetV2	RMSprop	$1 \times 10^{-3}$	2,266,951
MobileNetV3L	RMSprop	$1 \times 10^{-3}$	3,003,079
InceptionV3	Adam	$1 \times 10^{-3}$	21,817,127
ResNet50	Adam	$1 \times 10^{-3}$	23,602,055
ResNet50V2	SGD	$1 \times 10^{-2}$	23,579,143
ResNet101	RMSprop	$1 \times 10^{-3}$	42,672,519
ResNet101V2	SGD	$1 \times 10^{-2}$	42,640,903
Xception	RMSprop	$1 \times 10^{-3}$	20,875,823
ConvNeXtTiny	Adam	$1 \times 10^{-3}$	27,825,511
ConvNeXtLarge	Adam	$1 \times 10^{-3}$	196,241,095
EfficientNetB0	Adam	$1 \times 10^{-3}$	4,058,538
EfficientNetV2B0	Adam	$1 \times 10^{-3}$	5,928,279
EfficientNetV2L	Adam	$1 \times 10^{-3}$	117,755,815
EfficientNetB7	Adam	$1 \times 10^{-3}$	64,115,614

Among the pre-trained models on the TrashNet dataset, ResNet-50 and ConvNeXt-Large achieve the highest accuracy with 93% accuracy. According to the loss values, ConvNeXtLarge has shown outstanding performance, recording the lowest loss of 18%. This shows that ConvNeXtLarge not only achieves high accuracy but also remains effective in minimizing prediction errors. To exploit the collective power of different models, we evaluated various model combinations based on their robust balance between accuracy and loss.

The accuracy values of the average and weighted average ensemble models for different model combinations are shown in Table 8. The averaging ensemble model achieved 94.9% accuracy with the DenseNet121, ResNet50, ConvNeXtLarge model combination, while the weighted average ensemble model further improved performance, increasing accuracy to 96%. A summary of the results on the TrashNet dataset is presented in Table 9 and all experimental results are provided in Table A1.

**Table 8.** Evaluation of the averaging and weighted average ensemble models' accuracy on the TrashNet dataset.

Models	Averaging Ensemble (%)	Weighted Average Ensemble (%)	Weights
DenseNet121, ResNet50, ConvNeXtLarge	94.9	96.0	0.33, 0.22, 0.44
DenseNet121, ResNet101, ConvNeXtLarge	93.7	96.0	0.35, 0.18, 0.47
ResNet50, ResNet101, ConvNeXtLarge	94.9	95.3	0.17, 0.33, 0.5
DenseNet121, MobileNetV3, ConvNeXtLarge	94.1	95.3	0.17, 0.33, 0.5
ResNet50, MobileNetV3, ConvNeXtLarge	94.1	95.3	0.34, 0.06, 0.6
ResNet101, EfficientNetB0, ConvNeXtLarge	93.3	94.5	0.25, 0.25, 0.5
EfficientNetB7, Xception, ConvNeXtLarge	91.3	94.1	0.17, 0.16, 0.67
ResNet50, EfficientNetB0, ConvNeXtTiny	91.7	94.1	0.62, 0.31, 0.07

**Table 9.** Summary of experimental results on the TrashNet dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DenseNet121	90.5	88.1	91.9	89.4
ResNet50	92.9	90.6	92.6	91.4
ConvNeXtLarge	92.9	90.6	92.9	91.5
Averaging ensemble	94.9	93.0	94.0	94.0
Weighted average ensemble	96.0	94.0	97.0	95.0

In the experiments on the TrashBox dataset, ConvNeXtLarge is the best performing model with 94%. However, InceptionV3 shows the lowest accuracy among the models, with 80%. According to the loss values, ConvNeXtLarge shows the lowest loss with 17%. This shows that ConvNeXtLarge is the most effective model on the TrashBox dataset with high accuracy and low loss value. The evaluation of ensemble methods for different model combinations in the TrashBox dataset is presented in Table 10. The ResNet50, ResNet101, ConvNeXtLarge model combination obtained the highest accuracy on the TrashBox dataset. The averaging ensemble model achieved 93.5% accuracy and the weighted average ensemble model achieved 95.8% accuracy. A summary of the results on the TrashBox dataset is given in Table 11 and all experimental results are presented in Table A2.

**Table 10.** Evaluation of the averaging and weighted average ensemble models' accuracy on the TrashBox dataset.

Models	Averaging Ensemble (%)	Weighted Average Ensemble (%)	Weights
ResNet50, ResNet101, ConvNeXtLarge	93.5	95.8	0.07, 0.33, 0.6
ResNet101, EfficientNetB0, ConvNeXtLarge	94.1	95.6	0.33, 0.09, 0.58
DenseNet121, ResNet101, ConvNeXtLarge	93.8	95.6	0.09, 0.33, 0.58
DenseNet121, MobileNetV3, ConvNeXtLarge	93.4	95.5	0.1, 0.2, 0.7
DenseNet121, ResNet50, ConvNeXtLarge	93.0	95.4	0.18, 0.18, 0.64
EfficientNetB7, Xception, ConvNeXtLarge	93.9	95.3	0.22, 0.22, 0.56
ResNet50, EfficientNetV2B0, ConvNeXtTiny	91.5	91.9	0.3, 0.35, 0.35
DenseNet121, DenseNet201, EfficientNetB0	88.6	89.7	0.11, 0.33, 0.56

**Table 11.** Summary of experimental results on the TrashBox dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet50	85.9	85.8	85.9	85.8
ResNet101	87.0	87.1	87.0	87.0
ConvNeXtLarge	94.9	95.0	94.8	94.9
Averaging ensemble	93.5	93.0	93.0	93.0
Weighted average ensemble	95.8	96.0	96.0	96.0

In the experiments on the Waste Pictures dataset, ConvNeXtLarge is the top-performing model with an accuracy of 97%. It also demonstrates the lowest loss value at 10%, indicating its effectiveness on this dataset. The evaluation of ensemble methods for different model combinations in the Waste Pictures dataset is given in Table 12. The average ensemble model achieved 97.2% accuracy with the ResNet50, EfficientNetB0 and ConvNeXtLarge model combination, while the weighted average ensemble model reached 98% accuracy. A summary of the results on the Waste Pictures dataset is provided in Table 13 and all experimental results are shown in Table A3.

**Table 12.** Evaluation of the averaging and weighted average ensemble models' accuracy on the Waste Pictures dataset.

Models	Averaging Ensemble (%)	Weighted Average Ensemble (%)	Weights
ResNet50, EfficientNetB0, ConvNeXtLarge	97.2	98.0	0.22, 0.22, 0.56
ResNet101, EfficientNetB0, ConvNeXtLarge	96.8	98.0	0.19, 0.25, 0.56
EfficientNetB0, EfficientNetV2B0, ConvNeXtLarge	97.3	97.9	0.25, 0.19, 0.56
EfficientNetV2B0, ResNet50, ConvNeXtLarge	97.0	97.9	0.27, 0.18, 0.54
EfficientNetB0, InceptionV3, ConvNeXtLarge	97.0	97.9	0.31, 0.07, 0.62
DenseNet121, ResNet101, ConvNeXtLarge	96.0	97.8	0.1, 0.2, 0.7
ResNet50, ResNet101, ConvNeXtLarge	96.1	97.7	0.15, 0.23, 0.62
ResNet101V2, InceptionV3, ConvNeXtLarge	95.5	97.7	0.13, 0.12, 0.75

On the Garbage Classification dataset, ConvNeXtLarge is the most effective model, with 99% accuracy, and EfficientNetV2B0 follows close behind with 98% accuracy. On this dataset, we combine the predictions of EfficientNetB7, Xception and ConvNeXtLarge to create ensemble models. The optimal weights for maximum accuracy are 0.2, 0.3, 0.6, respectively. The average ensemble model shows an accuracy of 98.6% and the weighted average ensemble model shows an accuracy of 99.1%. The evaluation of ensemble methods for different model combinations is given in Table 14. A summary of the results for the

Garbage Classification dataset is presented in Table 15, and all experimental results are given in Table A4.

**Table 13.** Summary of experimental results on the Waste Pictures dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet50	90.8	91.0	89.8	90.2
ConvNeXtLarge	97.2	97.4	96.9	97.1
EfficientNetB0	93.8	93.9	93.5	93.6
Averaging ensemble	97.2	97.0	96.0	97.0
Weighted average ensemble	98.0	98.0	98.0	98.0

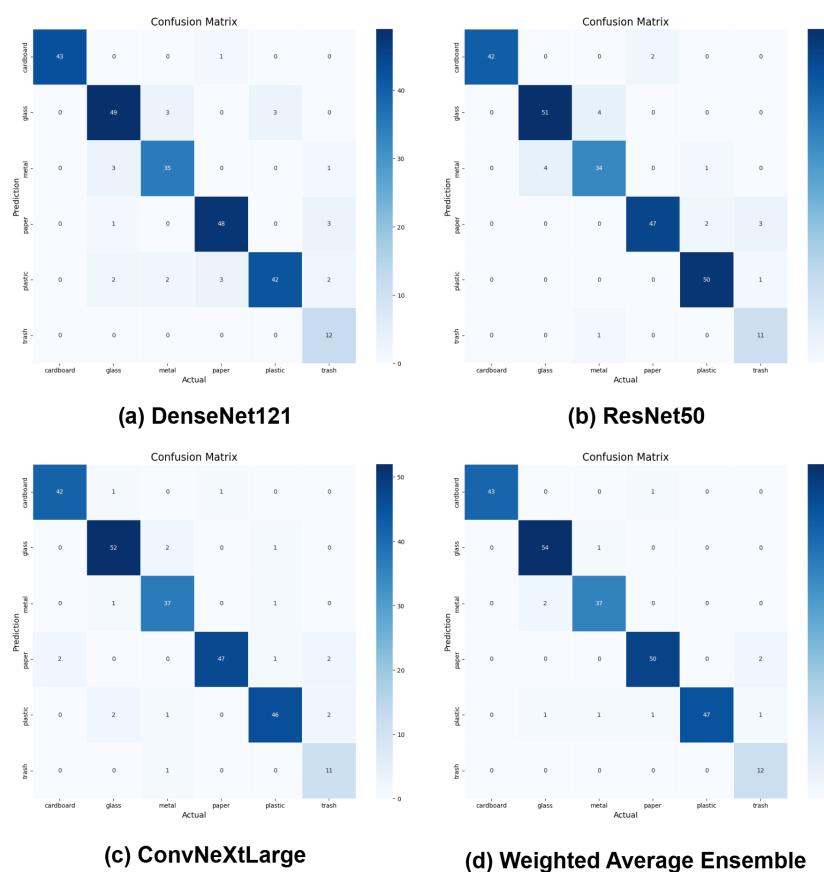
**Table 14.** Evaluation of the averaging and weighted average ensemble models' accuracy on the Garbage Classification dataset.

Models	Averaging Ensemble (%)	Weighted Average Ensemble (%)	Weights
EfficientNetB7, Xception, ConvNeXtLarge	98.6	99.1	0.2, 0.2, 0.6
DenseNet121, MobileNetV3, ConvNeXtLarge	98.0	99.0	0.06, 0.41, 0.53
ResNet50, ResNet101, ConvNeXtLarge	98.1	99.0	0.08, 0.25, 0.67
ResNet101, EfficientNetB0, ConvNeXtLarge	98.3	99.0	0.25, 0.25, 0.5
DenseNet121, ResNet101, ConvNeXtLarge	98.1	99.0	0.14, 0.14, 0.72
DenseNet121, ResNet50, ConvNeXtLarge	97.9	98.9	0.14, 0.14, 0.72
ResNet50, MobileNetV3, ConvNeXtLarge	98.3	98.9	0.2, 0.2, 0.6
ResNet50, EfficientNetV2B0, ConvNeXtTiny	97.6	97.7	0.1, 0.8, 0.1

**Table 15.** Summary of experimental results on the Garbage Classification dataset.

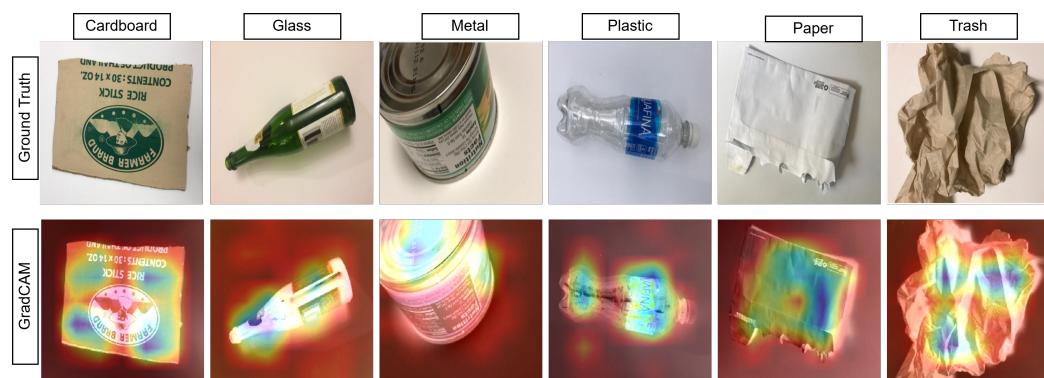
Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Xception	94.5	92.1	91.7	91.7
ConvNeXtLarge	98.8	98.6	98.7	98.6
EfficientNetB7	95.9	94.6	94.0	94.3
Averaging ensemble	98.6	98.0	98.0	98.0
Weighted average ensemble	99.1	99.0	99.0	99.0

Figure 5 shows the confusion matrix result of the selected base models and the weighted average ensemble on the TrashNet dataset. The weighted average ensemble demonstrates better performance in terms of classification accuracy and reduces misclassifications between classes with similar visual appearances. For instance, the baseline models confused paper with cardboard, whereas the ensemble model provided a clearer distinction between these classes. Furthermore, the ensemble model reduced misclassification errors between glass and plastic that commonly occurred due to varying light. The results demonstrate the ensemble approach's effectiveness in combining the complementary strengths of individual models and correcting their systematic errors.



**Figure 5.** Confusion matrices for TrashNet dataset. (a) DenseNet121, (b) ResNet50, (c) ConvNeXtLarge and (d) weighted average ensemble.

To enhance the explainability of our models, we apply Gradient-Weighted Class Activation Mapping (Grad-CAM) [68] to visualize the regions of input images that contribute most to classification decisions. Figure 6 shows the Grad-CAM visualizations of waste images from the TrashNet dataset for the ConvNeXtLarge model. The results indicate that the model consistently focuses on the most relevant object regions such as the central parts of recyclable materials (e.g., plastic bottles or paper surfaces), rather than irrelevant background areas. This provides that the model has learned meaningful and discriminative features and, by providing insight into the model's decision-making process, it effectively selects the relevant features of the waste items, thus ensuring that the waste is correctly classified. These findings validate the accuracy of the learned representations and also provide practical insights for the potential use of the system in real-world waste-sorting scenarios where explainability is critical for user trust and system transparency.



**Figure 6.** Grad-CAM visualizations of the ConvNeXtLarge model on the TrashNet dataset.

The paired *t*-test is applied using the results obtained from 5 to fold cross-validation to compare the proposed weighted ensemble model with the pre-trained models that performed best on the TrashNet dataset. Table 16 gives the *p*-values of the *t*-tests performed and the statistical significance of the differences in the values of the performance metrics. All the *p*-values are less than the 0.05 significance level, this confirms that the improvements achieved by the weighted ensemble are statistically significant. These results demonstrate that the ensemble model consistently outperforms pre-trained models, thereby validating its robustness and generalization ability.

**Table 16.** Statistical significance results on the TrashNet dataset.

Model	Accuracy <i>p</i> -Value	F1-Score <i>p</i> -Value
DenseNet121	0.001	0.001
ResNet50	0.002	0.002
MobileNetV3Large	0.001	0.001
EfficientNetV2B0	0.002	0.001
ConvNeXtLarge	0.02	0.03

Experimental results obtained on four waste classification datasets illustrate the effectiveness of the proposed ensemble learning approach. Although the individual DL models, including ResNet101, ConvNeXtLarge, and EfficientNetB0, perform well, the ensemble methods achieve superior accuracy, precision, recall and F1 scores. In particular, the weighted average ensemble method that is optimized through an extensive grid search proves its effectiveness, yielding the best classification accuracy across all datasets. Overall, this study provides a more robust and efficient solution for the automated waste sorting problem by integrating model diversity and optimized ensemble weighting. In addition, it supports future applications in practical smart recycling environments.

#### 4.4. Time Complexity Analysis

The deep learning models evaluated in this study are trained under the same experimental conditions to ensure a fair comparison. In particular, the same input image resolution, batch size and data augmentation configuration settings are used for all models. Furthermore, GPU acceleration is enabled to increase computational efficiency. Early stopping is applied during the training, and the reported times represent the convergence time of the model rather than the maximum epoch limits. Therefore, large and optimized models such as EfficientNet and ConvNeXt achieved faster convergence than smaller models. This demonstrates that modern architectures usually contain advanced optimization mechanisms and more efficient computational blocks, reaching convergence in less time with fewer epochs. Moreover, the training time analysis refers solely to the duration of individual model training and ensemble inference time is not included in this training time to avoid confusing the efficiency of model training with ensemble-level performance. The training time for all neural network models across four waste classification datasets is shown in Table 17.

The proposed ensemble method has a longer training time and higher computations on four datasets. This shows that training multiple models together increases the time cost of the processes. Although this strategic combination to improve the performance of the model improves accuracy, it has a significant disadvantage in terms of time and resource cost. This limits the application of this method, especially in systems with limited hardware resources or in real-time applications. Due to the dramatic increase in training time, maintaining, updating and retraining the model becomes more costly. Therefore, the applicability of the ensemble method should be carefully evaluated depending on the

usage scenario. This cost may be acceptable if the target is an academic or industrial project requiring high accuracy, but for resource-constrained systems, simpler and optimized models should be preferred.

**Table 17.** Training time of models.

Dataset	TrashNet	TrashBox	Waste Pictures	Garbage Classification
Model	Training Time(s)			
DenseNet121	770	3596	2842	2173
DenseNet201	674	3615	1254	2063
MobileNetV2	515	3582	1916	1997
MobileNetV3L	1001	4685	2377	1400
InceptionV3	746	3902	2165	2607
ResNet50	688	4600	1648	3408
ResNet50V2	737	4245	2122	1685
ResNet101	624	4585	2121	1399
ResNet101V2	842	3527	1670	1401
Xception	596	3556	2385	1702
ConvNeXtTiny	1034	5356	2840	3821
ConvNeXtLarge	429	4672	1483	2667
EfficientNetB0	497	2864	2845	1869
EfficientNetB7	1066	8689	1809	2585
EfficientNetV2B0	858	6033	3066	1860
EfficientNetV2L	882	7956	3750	3405
Proposed Method	1887	13,856	5975	6953

#### 4.5. Discussion

The combination of transfer learning and ensemble learning allows us to overcome the challenges of standalone ML/DL models. Transfer learning uses knowledge from a wider domain to improve the generalization capability of the model. This becomes specifically effective in using the proposed model in areas that have sparse or very inconsistent training data.

The proposed model makes use of pre-trained deep learning models to generalize from large datasets. This offers high accuracy and efficiency in addition to reduced training time. Moreover, because here the model's feature space depends upon diverse pre-trained data and not just the subtleties of the target data, this generalization in transfer learning will reduce overfitting on the target dataset and yield more reliable predictions. Ensemble learning combines several individual classifier predictions to improve the classification performance. Thus, it minimizes the weaknesses of individual models, increases the stability of the prediction and provides higher accuracy and robustness compared to using a single model. This approach is particularly useful in waste classification, where variability in waste types can create significant challenges. The experimental results clearly demonstrate that the presented ensemble methods obtain superior accuracy by better generalizing across different waste categories. Experimental results from existing studies in the literature are shown in Tables 18 and 19. Our results show improvement compared to previous studies; however, differences in data set splits, pre-processing steps and evaluation protocols between studies may affect the reported results. Therefore, the performance comparisons should be interpreted with these differences in mind.

**Table 18.** Performance comparison of different methods on TrashNet dataset.

Reference	Method	Train / Val / Test Ratio (%)	Accuracy (%)
(Yang et al., 2016) [20]	SIFT + SVM	70/13/17	63
(Kumsetty et al., 2022) [39]	Quantum ResNet-50	Not specified	80.5
(Bircanoglu et al., 2018) [21]	RecycleNet	70/13/17	81
(Adedeji et al., 2019) [31]	ResNet-50 + SVM	Not specified	87
(Rabano et al., 2018) [28]	MobileNet	Not specified	87.2
(Endah et al., 2020) [33]	Xception	80/20	88
(Ruiz et al., 2019) [32]	Inception-ResNet	80/10/10	88.6
(Satvikar et al., 2018) [29]	CNN	75/25	89.8
(Quan et al., 2024) [50]	VGG19	80/20	90.0
(Huang et al., 2023) [45]	ResNet18	80/20	91.4
(Azis et al., 2020) [35]	Inception-v3	80/10/10	92.5
(Shi et al., 2021) [36]	MLH-CNN,	80/20	92.6
(Kumsetty et al., 2023) [47]	ResNet-34	80/10/10	93.1
(Lin et al., 2024) [52]	VGG16	80/20	94.1
(Hossen et al., 2024) [48]	DenseNet201, MobileNet-v2	70/20/10	95.0
Proposed model (average ensemble)	DenseNet121, ResNet50, ConvNeXtLarge	80/10/10	94.9
Proposed model (weighted average ensemble)	DenseNet121, ResNet50, ConvNeXtLarge	80/10/10	96.0

**Table 19.** Performance comparison of different methods on Garbage Classification dataset.

Reference	Method	#Class	Train/Val/Test Ratio (%)	Accuracy (%)
(Chen et al., 2021) [42]	InceptionV3	12	80/10/10	93.1
(Shukurov, 2023) [43]	ResNeXt	12	80/10/10	95
(Dey et al., 2023) [44]	custom CNN	8	80/20	97.58
Proposed model (average ensemble)	EfficientNetB7, Xception, ConvNeXtLarge	12	80/10/10	98.6
Proposed model (weighted average ensemble)	EfficientNetB7, Xception, ConvNeXtLarge	12	80/10/10	99.1

The proposed weighted ensemble method outperforms state-of-the-art DL methods in waste classification. Most of the previous studies which are usually based on a single CNN architecture reported classification accuracies ranging from 80% to 95%, and often struggled with visually similar waste types, such as glass and plastic, or paper and cardboard. For instance, Endah et al. [33] proposed a method for waste classification using the Xception, VGG16 and ResNet-50 models on the TrashNet dataset. They obtained 88% accuracy with the Xception model. Azis et al. [35] introduced a DL-based method for waste classification using Inception-v3. The proposed model achieved an accuracy rate of 92.5% on TrashNet. Kumsetty et al. [47] achieved 93.1% accuracy on the TrashNet datasets using the ResNet34 model. Hossen et al. [48] presented RWC-Net which is based on a combination of DenseNet201 and MobileNet-V2 for waste classification. This model utilizes the feature extraction and learning capabilities of both pre-trained models. The proposed model achieved an accuracy rate of 95.01% on the TrashNet dataset. On the other hand, our ensemble approach leverages the complementary strengths of several architectures to achieve up to 96% accuracy. The proposed model provides better accuracy and efficiency for waste classification. The correct classification of waste is crucial for environmental sustainability and practical applications in waste management systems. Accurately and reliably classifying waste enables more efficient material recovery by minimizing the risks and costs of manual sorting. The proposed ensemble method is scalable and adaptable for real-world applications in waste management. The scalability of the proposed method allows for a number of integrations, such as AI-enabled collection bins, automated facility conveyors and smart city waste management infrastructure. Moreover, it can be deployable in distributed computing environments like cloud-based platforms or edge devices, enabling large-scale, real-time processing of waste images collected from multiple sources. It also contributes to global sustainability goals and environmental protection initiatives.

The results of this study emphasize the effectiveness of the ensemble methods, which show a significant improvement over individual models. Ensemble methods increase the

robustness and accuracy of the prediction model by combining different architectures, while individual models are capable on their own, they often suffer from limitations inherent to their specific architecture. For instance, a model can be quite effective in recognizing specific patterns but cannot generalize to a larger dataset due to overfitting. In contrast, ensemble methods alleviate these limitations by leveraging the strengths of multiple models. Ensemble methods reduce the variance in the predictions, thus resulting in a more reliable and accurate model. This research shows that the combination of different models within an ensemble framework can significantly improve overall performance in waste classification. Achieving high accuracy in waste classification is essential for effective waste management and recycling processes. The accurate classification of various waste types facilitates more efficient sorting processes, reduces pollution, and improves the quality of recycled materials.

While the proposed ensemble model provides an increase in accuracy compared to a single model, it comes with a significant computational cost. Training an ensemble model consists of running multiple models simultaneously, which not only increases training time but also increases inference time. This is an important practical consideration in real-world waste management systems where both efficiency and accuracy are important. In resource-constrained waste sorting environments, such as low-power IoT devices installed in recycling facilities, the inference latency is a major issue. For example, in conveyor belt-based waste sorting platforms, materials are transported at high speeds, and the classifier must provide real-time predictions to ensure accurate sorting. In this case, accuracy improvement cannot compensate for the increase in computational requirements, because system efficiency and computational costs are prioritized. Hence, utilizing a single fine-tuned model having optimal hyperparameters might be the most cost-effective solution in such systems. On the other hand, ensemble models can be useful in scenarios where the decisions are less time-sensitive but require high reliability. For instance, in offline applications like municipal waste management or environmental research, a little longer processing time is acceptable if the resulting classifications are more robust and less error-prone. Similarly, in high-risk applications such as hazardous or medical waste detection, the ensemble method can be favored since its more reliable predictions.

Although ensemble learning and transfer learning employing pre-trained models are well-known methods in computer vision, our work advances the field in meaningful ways, especially adapted to a challenging domain like waste classification. Our work specifically applies ensemble methods to waste classification which is a domain characterized by high intra-class variability, and noisy and real-world image conditions. We validate the proposed models on four different datasets, each with a different number of classes, image sources and environmental complexity. This detailed and comprehensive evaluation in our study has not been addressed in previous research. By evaluating multiple datasets of different complexity, our study provides a strong benchmark for future research in waste classification. Furthermore, our study compares two ensemble methods (averaging and weighted averaging) when combining pre-trained models on waste datasets, while previous studies generally use a single pre-trained model or simple CNN architectures, our work uniquely explores ensemble learning with many pre-trained models.

#### 4.6. Limitations and Future Work

While the proposed model has been evaluated on various datasets, there are still some limitations in their full applicability to real-world waste management environments. Although datasets such as TrashBox, Garbage Classification and Waste Pictures offer a wide variety of waste types and visual conditions due to being collected from different sources, they may lack a consistent representation of real-world waste. Furthermore, these

datasets contain images derived from online searches, which may introduce biases in composition, lighting and background that do not accurately reflect the unpredictable conditions encountered in waste bins, public spaces or industrial sorting facilities. The TrashNet dataset, while useful for controlled benchmarking, it is limited in both size and environmental variability. Although the datasets utilized in the study provide a baseline for benchmarking, they still lack some significant characteristics of real-world waste. In particular, the waste items in practical waste sorting scenarios are often overlapped or partially covered and this may mask key features and increase intra-class variability. Since most objects, like plastics, contaminated with food residues or fluids are dirty or soiled, they can distort the color, texture and shape information on which the model relies. Furthermore, inconsistent lighting conditions in waste sorting facilities and the mixed materials in the background can further complicate the image, thus making it difficult for models to generalize. These missing features in the datasets can lead to decreased accuracy, increased misclassification rates and reduced robustness when models are used in real-world waste management systems. Therefore, in future research we will evaluate our model with real waste datasets to make it more reliable and accurate in real-world applications. In order to further validate the robustness of our proposed ensemble learning approach, in future research work we will evaluate model performance on unseen, real-world data obtained from waste management systems. Real-world data will include all variability and uncertainty inherent in waste and allow for more intense testing of generalizability. Furthermore, methods such as data augmentation and synthetic data production can be used to address overfitting, especially where there are small or skewed datasets.

Real-world waste images commonly contain various types of noise, such as lighting variations, motion blur and background complexity. This can reduce the performance of DL models. Although the dataset used in this study includes clean images, dealing with noise is crucial to improve the robustness and scalability of waste classification systems used in uncontrolled environments. To address these difficulties, noise reduction techniques such as Gaussian filtering, median filtering and contrast normalization can be applied during the pre-processing stages. Recent studies [69,70] have shown the effect of noise on model accuracy and the potential for integrating noise reduction mechanisms into the classification process. These methods can be included in ensemble classification methods to increase classification accuracy in noisy environments. Future work will focus on integrating noise-sensitive pre-processing methods and noise reduction modules to improve the efficiency of the proposed ensemble method in naturally noisy environments.

A key limitation of our proposed deep ensemble learning method is the lack of interpretability common to most DL and ensemble models, while the methods generally obtain high accuracy, their complex architecture makes it challenging to provide clear, intuitive explanations for individual predictions. This “black box” nature can prevent confidence and transparency which are important factors for practical application in waste classification systems, especially when decisions can impact operational workflows or environmental compliance. As a result, users and other stakeholders may struggle to understand the rationale behind classification results and this can potentially limit adoption in environments where explainability is critical. In order to address this limitation, we used Grad-CAM [68] in this study. Moreover, post hoc explainability methods such as SHAP [71] and LIME [72] can also be used to help interpret model decisions.

Another limitation of our study is that while the proposed method has shown improved accuracy on various datasets, it is designed on established architectures rather than novel architectural designs or new learning algorithms that can better extract characteristic features of waste images. In our future work, we intend to create novel architectures or

custom feature extraction approaches specifically taking into account the challenges faced in waste classification.

We used basic averaging and weighted averaging techniques to combine the outputs of multiple pre-trained models. The reason for choosing these techniques was computational efficiency and ease of implementation which are essential properties for practical deployment in waste management systems that have limited computational resources, while these methods effectively improve the classification accuracy by exploiting the complementary strengths of individual models, there is still a need to develop advanced and potentially novel ensemble approaches for waste classification. Therefore, in our future work, we aim to explore some of the techniques such as dynamic weighting schemes that learn to adaptively assign weights based on model confidence or input features, stacking or meta-learning methods and designing custom ensemble architectures for waste classification. Future work exploring such novel approaches is valuable for improving the accuracy and efficiency of waste classification and improving the practical utility of these systems in various real-world scenarios.

## 5. Conclusions

This paper presents an effective deep ensemble learning model-based waste classification for sustainable waste management systems. The study utilizes high-performance, state-of-the-art CNN architectures and implements average and weighted average ensemble techniques to improve classification accuracy and model robustness. Furthermore, model explainability is ensured by using Grad-CAM for the waste classification task and the robustness and reliability of the findings are verified through five-fold cross-validation and paired *t*-tests. Experimental results on four different datasets have shown that the weighted average ensemble method outperforms the individual models. The weighted ensemble method performs at 96% accuracy, 94% precision, 97% recall, and 95% F1 score on the TrashNet dataset. These results clearly demonstrate that utilizing the complementary strengths of multiple CNN's improves the classification accuracy and also reduces the error rate in visually similar classes.

Although the study has achieved promising results, it has several limitations. First, the datasets used are of relatively small size and representation, hence may lack generalization to broader waste categories and real-world scenarios, including variable lighting conditions, surface contamination and deformation. Second, the performance of models is evaluated in a controlled setting rather than a real waste-sorting environment and is not evaluated in terms of hardware constraints like energy efficiency. In future research, large and diverse datasets will be included to enhance the robustness of the model. Moreover, real-time application studies using edge devices or conveyor-based sorting systems will assist in evaluating operational feasibility.

On the whole, the proposed ensemble model improves resource recovery, reduces landfill usage and protects the environment by providing an effective solution for improving automated recycling operations. With appropriate optimization and system integration, this approach can be widely applied in smart waste management infrastructures and also it can support circular economy initiatives and global sustainability goals.

**Author Contributions:** Conceptualization, A.A., F.Y.O. and İ.K.; methodology, A.A., F.Y.O. and İ.K.; software, A.A.; writing—original draft preparation, A.A.; writing—review and editing, A.A., F.Y.O., İ.K. and S.Ö; visualization, A.A.; supervision, S.Ö. All authors have read and agreed to the published version of the manuscript.

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

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Datasets used in this study are publicly available on Internet.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

**Table A1.** Experimental results on the TrashNet dataset.

Model	Macro Average				Weighted Average		
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Precision (%)	Recall (%)	F1-Score (%)
DenseNet121	90.5	88.1	91.9	89.4	91.2	90.5	90.6
DenseNet201	87.4	86.3	86.7	86.2	88.1	87.4	87.4
MobileNetV2	84.6	83.4	85.0	83.9	85.2	84.6	84.7
MobileNetV3L	89.3	87.1	89.9	88.2	89.8	89.3	89.4
InceptionV3	83.4	80.9	80.4	80.5	83.7	83.4	83.4
ResNet50	92.9	90.6	92.6	91.4	93.2	92.9	93.0
ResNet50V2	83.0	81.0	80.7	80.7	83.6	83.0	83.2
ResNet101	92.5	90.0	92.6	90.9	93.1	92.5	92.6
ResNet101V2	87.0	85.2	87.7	85.7	88.2	87.0	87.0
Xception	83.8	81.2	84.4	81.9	85.3	83.8	84.1
ConvNeXtTiny	84.6	82.4	84.6	82.7	86.1	84.6	84.8
ConvNeXtLarge	92.9	90.6	92.9	91.5	93.2	92.9	93.0
EfficientNetB7	86.6	83.2	86.5	84.0	87.9	86.6	86.8
EfficientNetB0	88.5	85.3	89.2	86.3	90.0	88.5	88.9
EfficientNetV2B0	89.3	86.3	89.9	87.3	90.6	89.3	89.6
EfficientNetV2L	85.8	81.6	83.6	81.9	87.7	85.8	86.3
Averaging ensemble	94.9	93.0	94.0	94.0	95.0	95.0	95.0
Weighted average ensemble	96.0	94.0	97.0	95.0	96.0	96.0	96.0

**Table A2.** Experimental results on the TrashBox dataset.

Model	Macro Average				Weighted Average		
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Precision (%)	Recall (%)	F1-Score (%)
DenseNet121	83.7	83.5	83.8	83.5	84.0	83.7	83.7
DenseNet201	85.1	85.1	85.3	84.9	85.8	85.1	85.2
MobileNetV2	80.6	80.5	80.7	80.2	81.1	80.6	80.5
MobileNetV3L	85.8	85.7	85.7	85.6	85.9	85.8	85.7
InceptionV3	79.6	79.7	79.7	79.3	80.3	79.6	79.6
ResNet50	85.9	85.8	85.9	85.8	86.0	85.9	85.8
ResNet50V2	82.4	82.8	82.0	82.3	82.7	82.4	82.4
ResNet101	87.0	87.1	87.0	87.0	87.2	87.0	87.0
ResNet101V2	84.9	84.8	84.8	84.7	84.9	84.9	84.8
Xception	82.1	82.3	82.0	82.0	82.5	82.1	82.2
ConvNeXt	87.2	87.1	87.2	87.1	87.3	87.2	87.2
ConvNeXtLarge	94.9	95.0	94.8	94.9	94.9	94.9	94.9
EfficientNetB7	87.3	87.1	87.4	87.2	87.4	87.3	87.3
EfficientNetB0	87.1	87.0	87.2	87.0	87.3	87.1	87.1
EfficientNetV2B0	88.3	88.0	88.3	88.1	88.4	88.3	88.3
EfficientNetV2L	87.5	87.3	87.5	87.4	87.5	87.5	87.5
Averaging ensemble	93.5	93.0	93.0	93.0	93.0	93.0	93.0
Weighted average ensemble	95.8	96.0	96.0	96.0	96.0	96.0	96.0

**Table A3.** Experimental results on the Waste Pictures dataset.

Model	Macro Average				Weighted Average		
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Precision (%)	Recall (%)	F1-Score (%)
DenseNet121	89.8	90.6	87.6	88.6	90.1	89.8	89.6
DenseNet201	90.3	90.8	88.9	89.5	90.6	90.3	90.3
MobileNetV2	89.3	89.9	88.7	89.0	89.7	89.3	89.3
MobileNetV3L	91.1	91.3	89.8	90.4	91.2	91.1	91.0
InceptionV3	86.5	87.3	84.8	85.6	87.0	86.5	86.5
ResNet50	90.8	91.0	89.8	90.2	90.9	90.8	90.7
ResNet50V2	89.7	90.8	88.4	89.2	90.2	89.7	89.7
ResNet101	91.8	91.9	90.3	90.9	92.0	91.8	91.8
ResNet101V2	88.0	87.8	86.7	86.9	88.5	88.0	87.9

**Table A3.** Cont.

Model	Macro Average				Weighted Average		
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Precision (%)	Recall (%)	F1-Score (%)
Xception	87.8	88.4	85.5	86.5	88.1	87.8	87.7
ConvNeXt	90.5	91.1	89.1	89.8	90.8	90.5	90.5
ConvNeXtLarge	97.2	97.4	96.9	97.1	97.2	97.2	97.2
EfficientNetB7	91.4	91.1	90.2	90.5	91.5	91.4	91.3
EfficientNetB0	93.8	93.9	93.5	93.6	93.8	93.8	93.8
EfficientNetV2B0	93.8	93.9	93.1	93.3	93.9	93.8	93.7
EfficientNetV2L	90.4	90.9	89.4	90.0	90.5	90.4	90.4
Averaging ensemble	97.2	97.0	96.0	97.0	97.0	97.0	97.0
Weighted average ensemble	98.0	98.0	98.0	98.0	98.0	98.0	98.0

**Table A4.** Experimental results on the Garbage Classification dataset.

Model	Macro Average				Weighted Average		
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Precision (%)	Recall (%)	F1-Score (%)
DenseNet121	95.4	93.9	93.2	93.5	95.5	95.4	95.4
DenseNet201	96.3	94.9	94.7	94.7	96.5	96.3	96.3
MobileNetV2	95.3	93.8	93.3	93.4	95.5	95.3	95.3
MobileNetV3L	95.6	94.2	93.7	93.9	95.7	95.6	95.6
InceptionV3	93.0	90.1	89.5	89.7	93.1	93.0	93.0
ResNet50	96.1	94.8	94.8	94.7	96.2	96.1	96.1
ResNet50V2	94.1	91.0	90.8	90.7	94.2	94.1	94.1
ResNet101	95.9	94.5	94.0	94.0	96.2	95.9	95.9
ResNet101V2	94.8	92.3	92.4	92.3	94.9	94.8	94.8
Xception	94.5	92.1	91.7	91.7	94.7	94.5	94.5
ConvNeXt	96.8	95.6	95.4	95.4	96.9	96.8	96.8
ConvNeXtLarge	98.8	98.6	98.7	98.6	98.9	98.8	98.8
EfficientNetB7	95.9	94.6	94.0	94.3	95.9	95.9	95.9
EfficientNetB0	96.3	95.0	94.9	94.9	96.4	96.3	96.3
EfficientNetV2B0	97.6	96.5	96.6	96.5	97.7	97.6	97.6
EfficientNetV2L	96.5	95.4	95.1	95.2	96.6	96.5	96.5
Averaging ensemble	98.6	98.0	98.0	98.0	99.0	99.0	99.0
Weighted average ensemble	99.1	99.0	99.0	99.0	99.0	99.0	99.0

## References

- Kaza, S.; Yao, L.; Bhada-Tata, P.; Van Woerden, F. *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*; World Bank Publications: Washington, DC, USA, 2018.
- Kang, Z.; Yang, J.; Li, G.; Zhang, Z. An Automatic Garbage Classification System Based on Deep Learning. *IEEE Access* **2020**, *8*, 140019–140029. [[CrossRef](#)]
- Huang, G.L.; He, J.; Xu, Z.; Huang, G. A combination model based on transfer learning for waste classification. *Concurr. Comput. Pract. Exp.* **2020**, *32*, e5751. [[CrossRef](#)]
- Yang, J.; Zeng, Z.; Wang, K.; Zou, H.; Xie, L. GarbageNet: A Unified Learning Framework for Robust Garbage Classification. *IEEE Trans. Artif. Intell.* **2021**, *2*, 372–380. [[CrossRef](#)]
- Okoya, S.A.; Oyinlola, M.; Schröder, P.; Kolade, O.; Abolfathi, S. Enhancing decentralised recycling solutions with digital technologies. In *Digital Innovations for a Circular Plastic Economy in Africa*; Taylor & Francis: Abingdon, UK, 2023; Volume 208.
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Pereira, F., Burges, C., Bottou, L., Weinberger, K., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; Volume 25.
- Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
- Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; PMLR: Cambridge MA, USA, 2015; pp. 448–456.
- Xie, S.; Girshick, R.; Dollar, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
- Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

12. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; Chaudhuri, K., Salakhutdinov, R., Eds.; PMLR: Cambridge MA, USA, 2019; Volume 97, pp. 6105–6114.
13. Pacal, I.; Kilicarslan, S.; Ozdemir, B.; Deveci, M.; Kadry, S. Efficient and autonomous detection of olive leaf diseases using AI-enhanced MetaFormer. *Artif. Intell. Rev.* **2025**, *58*, 303. [[CrossRef](#)]
14. Pacal, I.; Akhan, O.; Deveci, R.T.; Deveci, M. NeXtBrain: Combining local and global feature learning for brain tumor classification. *Brain Res.* **2025**, *1863*, 149762. [[CrossRef](#)] [[PubMed](#)]
15. Pacal, I.; Attallah, O. InceptionNeXt-Transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis. *Biomed. Signal Process. Control* **2025**, *110*, 108116. [[CrossRef](#)]
16. Pacal, I.; Attallah, O. Hybrid deep learning model for automated colorectal cancer detection using local and global feature extraction. *Knowl.-Based Syst.* **2025**, *319*, 113625. [[CrossRef](#)]
17. Nahavandi, S.; Alizadehsani, R.; Nahavandi, D.; Mohamed, S.; Mohajer, N.; Rokonuzzaman, M.; Hossain, I. A comprehensive review on autonomous navigation. *ACM Comput. Surv.* **2025**, *57*, 1–67. [[CrossRef](#)]
18. Cerrato, S.; Mazzia, V.; Salvetti, F.; Martini, M.; Angarano, S.; Navone, A.; Chiaberge, M. A deep learning driven algorithmic pipeline for autonomous navigation in row-based crops. *IEEE Access* **2024**, *12*, 138306–138318. [[CrossRef](#)]
19. Sakr, G.E.; Mokbel, M.; Darwich, A.; Khneisser, M.N.; Hadi, A. Comparing deep learning and support vector machines for autonomous waste sorting. In Proceedings of the 2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET), Beirut, Lebanon, 2–4 November 2016; pp. 207–212. [[CrossRef](#)]
20. Yang, M.; Thung, G. *Classification of Trash for Recyclability Status*; Technical Report; CS229 Project Report; Stanford University: Stanford, CA, USA, 2016.
21. Bircanoglu, C.; Atay, M.; Beşer, F.; Genc, O.; Kızıralı, M.A. RecycleNet: Intelligent Waste Sorting Using Deep Neural Networks. In Proceedings of the 2018 Innovations in Intelligent Systems and Applications (INISTA), Thessaloniki, Greece, 3–5 July 2018; pp. 1–7. [[CrossRef](#)]
22. Zeng, M.; Lu, X.; Xu, W.; Zhou, T.; Liu, Y. PublicGarbageNet: A Deep Learning Framework for Public Garbage Classification. In Proceedings of the 2020 39th Chinese Control Conference (CCC), Shenyang, China, 27–29 July 2020; pp. 7200–7205. [[CrossRef](#)]
23. Chen, Z.; Yang, J.; Chen, L.; Jiao, H. Garbage classification system based on improved ShuffleNet v2. *Resour. Conserv. Recycl.* **2022**, *178*, 106090. [[CrossRef](#)]
24. Sagi, O.; Rokach, L. Ensemble learning: A survey. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1249. [[CrossRef](#)]
25. Zhihong, C.; Hebin, Z.; Yanbo, W.; Binyan, L.; Yu, L. A vision-based robotic grasping system using deep learning for garbage sorting. In Proceedings of the 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017; pp. 11223–11226. [[CrossRef](#)]
26. Bai, J.; Lian, S.; Liu, Z.; Wang, K.; Liu, D. Deep Learning Based Robot for Automatically Picking Up Garbage on the Grass. *IEEE Trans. Consum. Electron.* **2018**, *64*, 382–389. [[CrossRef](#)]
27. Chu, Y.; Huang, C.; Xie, X.; Tan, B.; Kamal, S.; Xiong, X. Multilayer Hybrid Deep-Learning Method for Waste Classification and Recycling. *Comput. Intell. Neurosci.* **2018**, *2018*, 5060857. [[CrossRef](#)]
28. Rabano, S.L.; Cabatuan, M.K.; Sybingco, E.; Dadios, E.P.; Calilung, E.J. Common Garbage Classification Using MobileNet. In Proceedings of the 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Baguio City, Philippines, 29 November–2 December 2018; pp. 1–4. [[CrossRef](#)]
29. Satvilkar, M. Image Based Trash Classification Using Machine Learning Algorithms for Recyclability Status. Master’s Thesis, National College of Ireland, Dublin, Ireland, 2018.
30. Zeng, D.; Zhang, S.; Chen, F.; Wang, Y. Multi-Scale CNN Based Garbage Detection of Airborne Hyperspectral Data. *IEEE Access* **2019**, *7*, 104514–104527. [[CrossRef](#)]
31. Adedeji, O.; Wang, Z. Intelligent Waste Classification System Using Deep Learning Convolutional Neural Network. *Procedia Manuf.* **2019**, *35*, 607–612. [[CrossRef](#)]
32. Ruiz, V.; Sánchez, Á.; Vélez, J.F.; Raducanu, B. Automatic Image-Based Waste Classification. In Proceedings of the From Bioinspired Systems and Biomedical Applications to Machine Learning, Almería, Spain, 3–7 June 2019; Springer International Publishing: Cham, Switzerland, 2019; pp. 422–431.
33. Rismiyati; Endah, S.N.; Khadijah; Shiddiq, I.N. Xception Architecture Transfer Learning for Garbage Classification. In Proceedings of the 2020 4th International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 10–11 November 2020; pp. 1–4. [[CrossRef](#)]
34. Toğaçar, M.; Ergen, B.; Cömert, Z. Waste classification using AutoEncoder network with integrated feature selection method in convolutional neural network models. *Measurement* **2020**, *153*, 107459. [[CrossRef](#)]

35. Azis, F.A.; Suhaimi, H.; Abas, E. Waste Classification using Convolutional Neural Network. In Proceedings of the 2020 2nd International Conference on Information Technology and Computer Communications, Kuala Lumpur, Malaysia, 12–14 August 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 9–13. [[CrossRef](#)]
36. Shi, C.; Tan, C.; Wang, T.; Wang, L. A Waste Classification Method Based on a Multilayer Hybrid Convolution Neural Network. *Appl. Sci.* **2021**, *11*, 8572. [[CrossRef](#)]
37. Zhang, Q.; Zhang, X.; Mu, X.; Wang, Z.; Tian, R.; Wang, X.; Liu, X. Recyclable waste image recognition based on deep learning. *Resour. Conserv. Recycl.* **2021**, *171*, 105636. [[CrossRef](#)]
38. Majchrowska, S.; Mikołajczyk, A.; Ferlin, M.; Klawikowska, Z.; Plantykov, M.A.; Kwasigroch, A.; Majek, K. Deep learning-based waste detection in natural and urban environments. *Waste Manag.* **2022**, *138*, 274–284. [[CrossRef](#)] [[PubMed](#)]
39. Kumsetty, N.V.; Bhat Nekkare, A.; Sowmya, S.K.; Kumar, A.M. TrashBox: Trash Detection and Classification using Quantum Transfer Learning. In Proceedings of the 2022 31st Conference of Open Innovations Association (FRUCT), Helsinki, Finland, 27–29 April 2022; pp. 125–130. [[CrossRef](#)]
40. Zhou, H.; Yu, X.; Alhaskawi, A.; Dong, Y.; Wang, Z.; Jin, Q.; Hu, X.; Liu, Z.; Kota, V.G.; Abdulla, M.H.A.H.; et al. A deep learning approach for medical waste classification. *Sci. Rep.* **2022**, *12*, 2159. [[CrossRef](#)]
41. Waste Pictures. 2019. Available online: <https://www.kaggle.com/wangziang/waste-pictures> (accessed on 1 May 2024).
42. Chen, Y.; Han, W.; Jin, J.; Wang, H.; Xing, Q.; Zhang, Y. Clean Our City: An Automatic Urban Garbage Classification Algorithm Using Computer Vision and Transfer Learning Technologies. *J. Phys. Conf. Ser.* **2021**, *1994*, 012022. [[CrossRef](#)]
43. Shukurov, R. Garbage classification based on fine-tuned state-of-the-art models. In Proceedings of the 2023 9th International Conference on Control, Decision and Information Technologies (CoDIT), Rome, Italy, 3–6 July 2023; pp. 841–846. [[CrossRef](#)]
44. Dey, D.; Shama, U.S.; Akash, M.; Karim, D.Z. Automatic Waste Classification System using Deep Learning Techniques. In Proceedings of the 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Tenerife, Canary Islands, Spain, 19–21 July 2023; pp. 1–6. [[CrossRef](#)]
45. Huang, L.; Li, M.; Xu, T.; Dong, S.Q. A waste classification method based on a capsule network. *Environ. Sci. Pollut. Res.* **2023**, *30*, 86454–86462. [[CrossRef](#)]
46. Jin, S.; Yang, Z.; Królczykg, G.; Liu, X.; Gardoni, P.; Li, Z. Garbage detection and classification using a new deep learning-based machine vision system as a tool for sustainable waste recycling. *Waste Manag.* **2023**, *162*, 123–130. [[CrossRef](#)]
47. Kumsetty, N.V.; Nekkare, A.B.; Sowmya Kamath, S.; Anand Kumar, M. An Approach for Waste Classification Using Data Augmentation and Transfer Learning Models. In *Machine Vision and Augmented Intelligence: Select Proceedings of MAI 2022*; Springer Nature: Singapore, 2023; pp. 357–368. [[CrossRef](#)]
48. Hossen, M.M.; Majid, M.E.; Kashem, S.B.A.; Khandakar, A.; Nashbat, M.; Ashraf, A.; Hasan-Zia, M.; Kunju, A.K.A.; Kabir, S.; Chowdhury, M.E.H. A Reliable and Robust Deep Learning Model for Effective Recyclable Waste Classification. *IEEE Access* **2024**, *12*, 13809–13821. [[CrossRef](#)]
49. Lilhore, U.K.; Simaiya, S.; Dalal, S.; Damaševičius, R. A smart waste classification model using hybrid CNN-LSTM with transfer learning for sustainable environment. *Multimed. Tools Appl.* **2024**, *83*, 29505–29529. [[CrossRef](#)]
50. Quan, M.K.; Nguyen, D.C.; Nguyen, V.D.; Wijayasundara, M.; Setunge, S.; Pathirana, P.N. Toward Privacy-Preserving Waste Classification in the Internet of Things. *IEEE Internet Things J.* **2024**, *11*, 24814–24830. [[CrossRef](#)]
51. Sarswat, P.K.; Singh, R.S.; Pathapati, S.V.S.H. Real time electronic-waste classification algorithms using the computer vision based on Convolutional Neural Network (CNN): Enhanced environmental incentives. *Resour. Conserv. Recycl.* **2024**, *207*, 107651. [[CrossRef](#)]
52. Lin, Z.; Xu, H.; Zhou, M.; Wang, B.; Qin, H. Waste classification strategy based on multi-scale feature fusion for intelligent waste recycling in office buildings. *Waste Manag.* **2024**, *190*, 443–454. [[CrossRef](#)]
53. Kumar, A.K.; Ali, Y.; Kumar, R.R.; Assaf, M.H.; Ilyas, S. Artificial Intelligent and Internet of Things framework for sustainable hazardous waste management in hospitals. *Waste Manag.* **2025**, *203*, 114816. [[CrossRef](#)]
54. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [[CrossRef](#)]
55. Donnelly, J.; Daneshkhah, A.; Abolfathi, S. Forecasting global climate drivers using Gaussian processes and convolutional autoencoders. *Eng. Appl. Artif. Intell.* **2024**, *128*, 107536. [[CrossRef](#)]
56. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.
57. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
58. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

59. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
60. Bhuiyan, M.; Islam, M.S. A new ensemble learning approach to detect malaria from microscopic red blood cell images. *Sensors Int.* **2023**, *4*, 100209. [CrossRef]
61. Vij, R.; Arora, S. A hybrid evolutionary weighted ensemble of deep transfer learning models for retinal vessel segmentation and diabetic retinopathy detection. *Comput. Electr. Eng.* **2024**, *115*, 109107. [CrossRef]
62. Kang, M.; Ahn, J.; Lee, K. Opinion mining using ensemble text hidden Markov models for text classification. *Expert Syst. Appl.* **2018**, *94*, 218–227. [CrossRef]
63. Chen, H.; Zhang, Z.; Huang, S.; Hu, J.; Ni, W.; Liu, J. TextCNN-based ensemble learning model for Japanese Text Multi-classification. *Comput. Electr. Eng.* **2023**, *109*, 108751. [CrossRef]
64. Cao, Y.; Wang, Z.; Ding, H.; Zhang, J.; Li, B. An intrusion detection system based on stacked ensemble learning for IoT network. *Comput. Electr. Eng.* **2023**, *110*, 108836. [CrossRef]
65. Jethanandani, M.; Sharma, A.; Perumal, T.; Chang, J.R. Multi-label classification based ensemble learning for human activity recognition in smart home. *Internet Things* **2020**, *12*, 100324. [CrossRef]
66. Thung, G.; Yang, M. Trashnet. 2016. Available online: <https://github.com/garythung/trashnet> (accessed on 1 May 2024).
67. Garbage Classification. 2021. Available online: <https://www.kaggle.com/datasets/mostafaabla/garbage-classification> (accessed on 1 May 2024).
68. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
69. Attallah, O.; Ibrahim, R.A.; Zakzouk, N.E. A lightweight deep learning framework for transformer fault diagnosis in smart grids using multiple scale CNN features. *Sci. Rep.* **2025**, *15*, 14505. [CrossRef] [PubMed]
70. Chen, C.; Liang, R.; Song, M.; Zhang, Z.; Tao, J.; Yan, B.; Cheng, Z.; Chen, G. Noise-assisted data enhancement promoting image classification of municipal solid waste. *Resour. Conserv. Recycl.* **2024**, *209*, 107790. [CrossRef]
71. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
72. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should i trust you?” Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.

**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 MDPI and/or the editor(s). MDPI 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.