SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

TriFusionNet: A Heterogeneous Feature Fusion Approach for Multi-Class Waste Image Classification

Ankush[1*], Dr Sunil Mankotia[2]

[1]Department of Computer Science, Himachal Pradesh University, Shimla, India. E-mail: ankushkabir6@gmail.com
[2] University College of Business Studies (UCBS), Avalodge, Shimla, India. E-mail: mankotia.sunil@gmail.com

Abstract

The growing waste volume and the variety of waste materials demand reliable, automatic waste classification systems. While deep learning techniques have been effective for waste image classification, existing methods are limited to a single feature extraction backbone, potentially missing the rich set of visual attributes of waste materials. To solve this problem, this paper introduces TriFusionNet, a three-branch heterogeneous feature-fusion network consisting of EfficientNet-B0, ResNet18 and a lightweight custom CNN. The framework extracts high-level, residual, and task-specific local feature representations from the input images and combines them into a unified 1920-dimensional feature representation for classification. The images are integrated from both the TrashNet and Waste Classification Data datasets to create a balanced 5-class waste dataset of 29,085 images, including glass, metal, organic, paper and plastic. The model is based on the idea of using pre-trained backbones with transfer learning and data augmentation, and regularization to enhance generalization. Accuracy, precision, recall, F1-score, confusion matrix and ROC-AUC are used to assess the effectiveness of the proposed framework. A systematic ablation study involving seven configurations is conducted to investigate the contribution of the individual and combined feature-extraction branches. The total classification accuracy of TriFusionNet is 97.16%, and the ablatio results show that the full three-branch configuration yields the highest accuracy with respect to the different configurations evaluated. In addition, the image regions that support the model's predictions are qualitatively analyzed using Grad-CAM visualization. The results show that the proposed heterogeneous feature fusion approach demonstrates the potential of heterogeneous feature fusion for robust five-class waste image classification.

Keywords: Waste classification, deep learning, heterogeneous feature fusion, transfer learning, EfficientNet-B0, ResNet-18, Custom CNN, Grad-CAM.

1. Introduction

Municipal solid waste (MSW) production, an outcome of urbanization, industrialization and population growth, is a significant challenge in environmental and socio-economic problems globally [1]. Recent environmental reports indicate that in an average year, there are millions of tons of waste produced, putting a tremendous strain on current waste management systems and lowering the effectiveness of recycling systems. Uncontrolled disposal of waste that can be recycled can cause environmental pollution, the greenhouse effect, depletion of resources and serious health issues. Therefore, smart waste management has become an integral part of sustainable urban development and circular economy projects [2]. Waste segregation is the first and one of the most important steps in the recycling process. The correct categorization of waste into the various categories of materials facilitates efficient recycling, reduces the mix of waste materials and increases the recovery of resources. But traditional sortation by hand is still the norm in many recycling centres [3]. Manual sorting is time-consuming, requires a lot of labor, is costly and is prone to human error due to worker fatigue, inconsistent sorting determination and the rising range of waste materials. These limitations make it necessary to create an automated waste categorization system with high accuracy, speed, and dependability [4]. In recent years, computer vision tasks including object identification, image recognition, semantic segmentation, and automated trash sorting have been transformed by artificial intelligence (AI) and deep learning (DL). Recently, Convolutional Neural Networks (CNNs) have shown great power in learning hierarchical visual representations automatically from raw images, with no need for manual feature engineering. Therefore, CNN-based methods are the most preferred approach for intelligent waste classification, as they are more accurate than traditional machine learning methods that rely on manually designed descriptors [5].

Waste classification using deep learning has been further gaining ground thanks to transfer learning. Pre-trained architectures like VGG, ResNet, DenseNet, MobileNet, EfficientNet, and Vision Transformers have been modified and successfully deployed to waste image datasets using knowledge gained from large-scale benchmark image datasets like ImageNet. The improvement in the classification accuracy of these models is remarkable, and the training time and computational requirements have been reduced significantly. These architectures- Compound Scaling in EfficientNet yield an efficient trade-off between model complexity and predictive performance, and ResNet uses residual learning to allow for training deeper neural networks and better feature representation [6]. While existing deep learning models have showcased promising results, there are still numerous challenges that need to be addressed. First, most studies have been based on a single backbone network, so the diversity of learned representations is limited, and the model may struggle to differentiate visually similar waste categories. Secondly, owing to variations in illumination, viewpoint, background clutter, deformation, occlusion and material texture, Recyclable waste has considerable inter-class commonality and intra-class heterogeneity. These features may make correct discrimination difficult for specific CNN designs. Third, lighter models tend to have less discriminative power to be more computationally efficient, while deeper models tend to need more computational power and so are more difficult to deploy in practical recycling systems. The limitations indicate that a single feature extraction network might not be able to capture all the complementary visual features present in heterogeneous waste images. To boost classification performance, researchers have tried various approaches, including the use of ensemble learning, transfer learning, attention mechanisms, and hybrid models. Most of the current hybrid methods, however, integrate deep feature extraction with conventional machine learning classifiers, or focus on enhancing a single backbone network. Compared to this, fewer studies focus on using more than one complementary deep neural network in a single, end-to-end feature fusion structure that can learn semantic, structural and local texture information all at once. Hence, there is still ample room for further developing stronger hybrid architectures that realize effective integration of complementary feature representations with high classification accuracy and good computational efficiency [7]. Unlike HES-Net, which employs two pretrained branches (EfficientNet-B0 and Shuffle) [8], the present work introduces a three-branch heterogeneous feature-fusion architecture by incorporating a task-specific custom CNN and ResNet18. The present study also evaluates all individual, pairwise, and three-branch configurations through a controlled seven-configuration ablation study. To tackle this problem, in this study, a hybrid deep learning architecture named TriFusionNet is proposed, which combines EfficientNet-B0, ResNet18, and a lightweight Custom CNN for waste classification, intelligently divide into multiple classes. The proposed architecture extracts high-level semantic representations via compound feature scaling in EfficientNet-B0, learns robust residual features preserving discriminative information through deep layers of the network in ResNet18, and learns fine-grained local texture features in the Custom CNN. The feature vectors from these 3 pathways are fused together into a high-dimensional vector, fed to fully connected classification layers, including batch normalization, dropout regularization and label smoothing, for improved generalization in the model. Furthermore, the transfer learning process is divided into two stages: freezing and fine-tuning the pretrained backbone networks are used to realize effective domain adaptation and anti-overfitting. A balanced dataset of waste is used to assess the proposed framework. Images of five different waste categories (glass, metal, organic, paper and plastic) were used. Paper and plastic. Images were compiled to form the dataset. From available public datasets, TrashNet and Waste Classification Data. A total of 29,085 images were employed, and 5,817 images were assigned to each class. The dataset was split into training, validation, and test sets. test sets contained 20,355, 4,360, and 4,370 images, respectively. The resulting performance of the proposed TriFusionNet was assessed with the aid of the following metrics: accuracy, precision, recall, F1-score, Use of confusion matrix analysis, and class-wise performance analysis. Three backbone networks are selected in the TriFusionNet which are not just for the sheer number of network branches, but for complementary representation characteristics. EfficientNet-B0 is used to extract compact high-level semantic representations using its efficient convolutional network architecture, and ResNet18 is used to extract discriminative feature representations using residual connections. The lightweight Custom CNN, on the other hand, is tailored to learn local visual and texture patterns that are relevant to the task and might not be learnt sufficiently by the pretrained backbones. The Custom CNN is thus here to support the more efficient EfficientNet-B0 branch, and not to replace it. The controlled ablation study, where single, paired, and three-branch configurations are compared under the same experimental conditions, corroborates the design decision. Importantly, the proposed framework does not require prediction-level averaging or weighted voting that is typical of ensemble approaches, but instead uses different representations of features in parallel and concatenates them prior to the classification head, for the purpose of jointly learning from semantic, residual and local representations of features. Hence, the main advantage of TriFusionNet is the controlled combination of three different heterogeneous feature representations in a single feature-level fusion method for five-class classification of waste. The major findings of this study are summarized as follows:

1. A three-branch heterogeneous feature-fusion framework, termed TriFusionNet, is proposed by integrating EfficientNet-B0, ResNet18, and a lightweight custom CNN to combine complementary high-level, residual, and task-specific local feature representations for five-class waste classification.
2. A feature-level fusion mechanism is developed in which the 1280-dimensional EfficientNet-B0, 512-dimensional ResNet18, and 128-dimensional custom CNN representations are concatenated to form a unified 1920-dimensional feature representation before classification.

3. A two-stage training strategy is employed for the pretrained EfficientNet-B0 and ResNet18 branches, together with batch normalization, dropout, label smoothing, learning-rate scheduling, gradient clipping, and early stopping to improve optimization stability and generalization.
4. The proposed framework was comprehensively evaluated on a 5-class waste image dataset, achieving an overall classification accuracy of 97.16%, along with consistently high precision, recall, and F1-score across all waste categories.
5. The proposed model is a promising and efficient way to tackle the challenge of intelligent waste recognition, with potential applications in automated recycling systems, smart waste management platforms, and IoT-based environmental monitoring systems.

2. Related Work

In recent years, automated waste classification and recycling systems have come a long way, greatly enhanced by computer vision and deep learning technologies. Most waste separation is done manually, which is time-consuming, labor-intensive and prone to human error. To overcome these limitations, image processing and deep learning methods have been increasingly applied to automate waste identification and classification. CNNs, which can learn automatically hierarchical and discriminative visual representations from images, are very effective.

CNN-based classification has been extensively studied in waste management and has proven useful in other image-based quality-assessment applications. Jahanbakhshi et al. [9] used a CNN-based approach to classify carrot images into normal and abnormal categories using 878 images. The proposed model achieved an accuracy of 99.43% after incorporating data augmentation, which shows that CNNs can learn the discriminative visual characteristics for automated quality assessment. This work is not related to waste material as the other works, but it provides proof of the effectiveness of visual classification using CNNs for automated inspection. But the classification of waste is not so straightforward since waste objects have significant differences in shape, texture, lightness, deformation and background conditions.

A. CNN-Based and Transfer-Learning Approaches for Waste Classification

Following the success of CNNs, the idea of transfer learning, that is, using a pre-trained architecture for better waste-classification performance, has been explored. Kaya et al. [10] tested the custom CNN against some popular CNN models such as VGG19, ResNet101, InceptionV3, Xception and DenseNet169 on the TrashNet dataset. The custom CNN attained an accuracy of 88.66%, while DenseNet169 had the highest accuracy of 96.42% and an F1 of 96%. The results show the ability of pretrained deep architectures to learn strong visual representations from images of waste. However, the study focused mainly on individual architectures and did not explore fusing complementary feature representations from multiple CNN backbones.

Also, there has been a lot of interest in lightweight architectures because real-world waste sorting systems may face computational and resource limitations. In a study by Lilhore et al. [11], the authors proposed a smart waste management system using CNN-based architectures, such as ResNet, Inception, Xception, and MobileNet, with cloud computing support. The model with the highest reported accuracy of 94.44% was achieved by MobileNetV3, which is a balance between computational efficiency and classification performance. However, relying on the cloud for processing could cause communication delays and could restrict where it can be deployed, such as in resource-limited edge environments. In addition, deep-learning models that do not explicitly include interpretability mechanisms may be hard to comprehend when it comes to understanding the visual evidence for specific predictions.

B. Hybrid and Multi-Architecture Deep-Learning Approaches

In this regard, recent works have investigated how to go beyond the scope of individual feature extractors and make use of hybrid deep-learning architectures that use complementary representation-learning architectures. To use the complementary deep features, Hossen et al. [12] suggested a hybrid structure that combined DenseNet201 and InceptionV3 backbones. The framework first separated the images into garbage and non-garbage classes and then distinguished the waste items such as bulky waste, cardboard, litter and garbage bags from the GIGO dataset. The authors also added Bidirectional Long Short-Term Memory (Bi-LSTM) to the transfer-learning paradigm to capture the relationship between the extracted features. The results showed that in complex waste management scenarios, CNN-based feature extraction along with Bi-LSTM resulted in better robustness in classification. However, the architecture focuses on the problem of hierarchical classification rather than the three-branch architecture of heterogeneous CNN representations for common categories of recyclable materials.

Analogous deep-learning approaches have been utilized in specific waste-management areas. Several traditional machine learning methods, such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN), have been studied for construction and demolition waste (C&DW). Most of these approaches, however, rely on hand-crafted representations of features, which may not be able to represent complex visual patterns. Davis et al. [13] showed that CNN-based techniques can enhance the accuracy of classification by learning visual representation hierarchies directly from images. For E-Waste recycling, Sarswat et al. [14] explored deep-learning-based object detection using the YOLOv5 and YOLOv7 networks for real-time applications. They achieved about 94% classification accuracy and about 0.96 Average Precision (AP) for YOLOv7, showcasing the promise of one-stage object-detection networks in real-world recycling tasks. Construction and electronic-waste studies, however, are focused on specific areas of waste, which do not directly provide solutions for the five-class material classification problem examined in this study.

C. Multi-Stage Classification for Complex Waste Categories

With the growing number and variety of waste types, hierarchical and multi-stage classification have also been studied. Nahiduzzaman et al. [15] have introduced a three-stage waste-classification scheme for electronic components which employs a lightweight depthwise separable CNN (DP-CNN) and Ensemble Extreme Learning Machine (En-ELM). The framework was tested in a binary classification mode, a nine-class mode and a 36-class mode, and it has been able to reach levels of approximately 96%, 91% and 85.25% accuracy, respectively. The performance drop with the number of categories shows that waste classification becomes more and more challenging the more categories there are. The hierarchical classification approach can be used to organize the complex category structure well but cannot directly solve the fusion of features from multiple complementary CNN backbones.

D. Alternative Sensing Techniques for Waste Identification

Besides image-based deep learning, other sensing technologies have been explored for the identification of materials. Das et al. [16] discussed spectroscopy-based methods such as IR, Raman, XRF and LIBS for the identification of polymers in e-waste. The study found an accuracy of around 90% using SVM, while KNN and SIMCA had an accuracy above 98% in a controlled laboratory environment. These techniques can, however, give very specific material information but require special sensing equipment, which can add system complexity and cost. Therefore, image-based methods are still appealing for smart waste-management systems that require low cost and scalability.

E. Research Gap and Motivation

Previous studies have shown that CNNs, transfer learning, lightweight architectures and hybrid models are effective for the classification of waste, but there are still some drawbacks. First, many approaches are based on a single feature-extraction backbone, which means that they learn representations mainly from a single architectural angle. Second, existing hybrid approaches do not consistently investigate feature-level integration of heterogeneous CNN representations for five-class material-based waste classification. Thirdly, waste categories that are similar in visual appearance can have significant intra-class variation and inter-class similarity due to differences in 'shape', 'texture', 'deformation', illumination', and 'background'. These properties encourage the adoption of complementary feature representations instead of single backbone representations. In order to overcome these drawbacks, we introduce TriFusionNet, a three-branch heterogeneous feature-fusion network composed of EfficientNet-B0, ResNet18 and a lightweight custom CNN. The proposed framework explicitly integrates complementary feature representations prior to classification, where EfficientNet-B0 is used to provide high-level feature representations, ResNet18 to provide complementary residual representations, and the custom CNN to capture task-specific local patterns. All three representations are then combined into a single, 1920-dimensional feature vector and passed through a shared classification head. To explore the contribution of individual and combined branches, a systematic ablation study of 7 configurations is additionally performed, and the Grad-CAM visualization is used to qualitatively analyze the regions that affect the model predictions.

3. PROPOSED TRIFUSIONNET FRAMEWORK

A. Overview of the Proposed Framework

TriFusionNet receives an RGB image of size $224 \times 224 \times 3$ and processes it through three parallel feature-extraction branches. EfficientNet-B0 produces a 1280-dimensional feature representation, ResNet18 produces a 512-dimensional feature representation, and the Custom CNN produces a 128-dimensional feature representation. These three feature vectors are concatenated to form a 1920-dimensional fused representation. The resulting feature vector is then passed through the classification head with the dimensions $1920 \rightarrow 512 \rightarrow 256 \rightarrow 5$.

EfficientNet-B0 and ResNet18 are initialized with pretrained weights. During the initial transfer-learning stage, their parameters are frozen to preserve the learned generic visual representations. The pretrained backbones are subsequently unfrozen for task-specific fine-tuning, whereas the Custom CNN is trained as part of the proposed hybrid architecture.

Figure 1 shows the overall architecture of the proposed TriFusionNet. The framework consists of the following major stages:

1. Dataset acquisition and preparation from the TrashNet (Garbage classification) and Waste Classification datasets.
2. Image preprocessing and data augmentation.
3. Parallel feature extraction using EfficientNet-B0, ResNet18, and the Custom CNN.
4. Feature fusion through concatenation of the extracted feature representations.
5. Classification using the fully connected classification head.

6. Model training, validation, and performance evaluation using standard classification metrics.

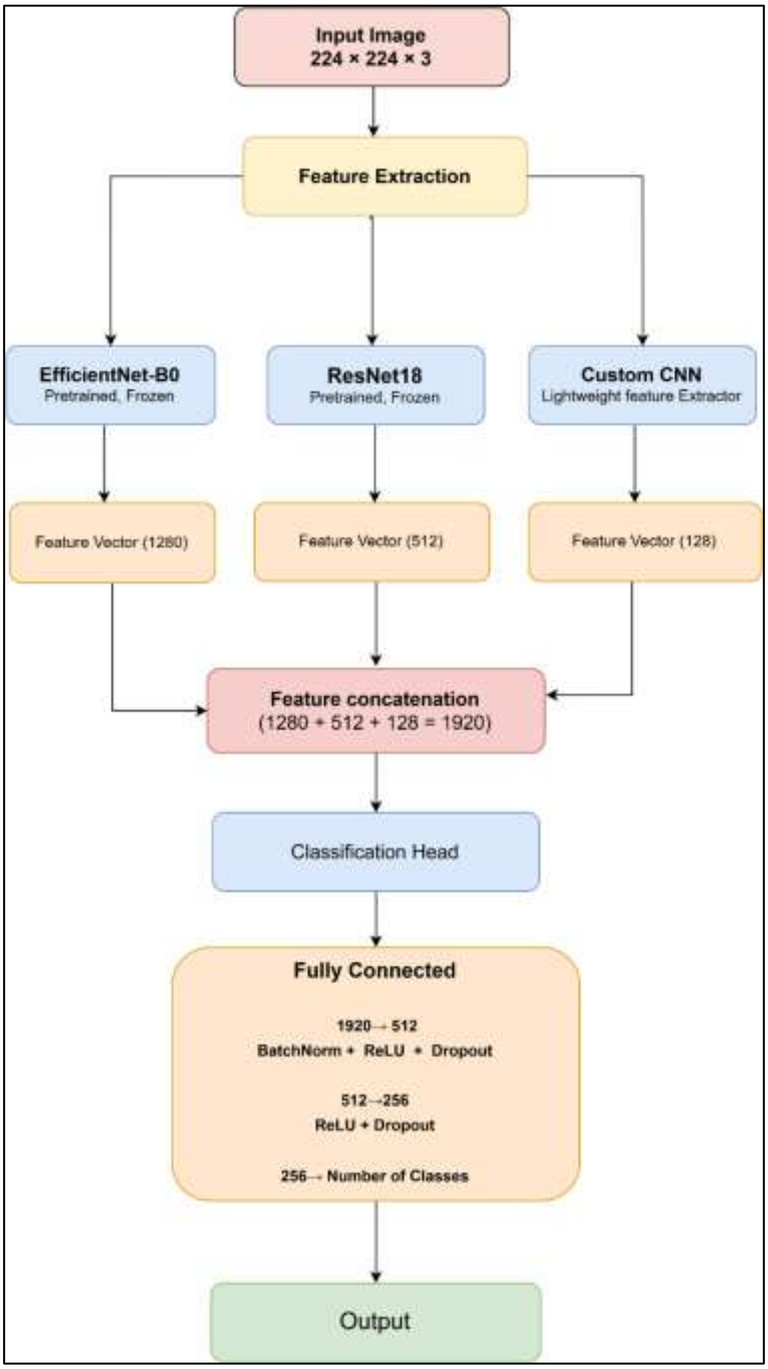

Fig.1 Proposed TriFusionNet framework for heterogeneous feature-level fusion of EfficientNet-B0, ResNet18, and a task-specific custom CNN for five-class waste classification.

B. Dataset Preparation

The proposed TriFusionNet was developed using a unified waste-image dataset constructed from the publicly available TrashNet (Garbage Classification) and Waste Classification Data datasets [17], [18]. To establish a consistent classification objective across the two datasets, five common waste categories- glass, metal, organic, paper, and plastic-were selected. The images from both datasets were integrated following preprocessing and quality verification to remove unsuitable samples and ensure consistency. The resulting dataset comprised 29,085 images, with 5,817 images per category, thereby maintaining an equal class distribution and minimizing the influence of class imbalance during model training and evaluation. All images were resized to 224 × 224 pixels and normalized using ImageNet statistics. In addition, random data augmentation techniques were applied to the training images to increase sample diversity, reduce overfitting, and improve the generalization capability of the proposed model.

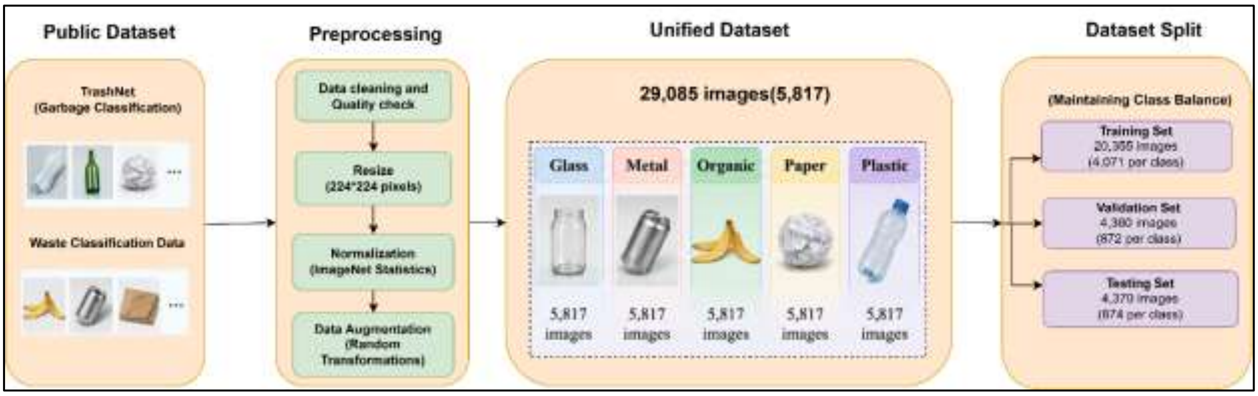

Fig.2: Dataset integration, preprocessing, class balancing, and data partitioning for waste classification.

The balanced dataset was divided into training, validation, and testing subsets while preserving the same class distribution across all three subsets. Specifically, each waste category contained 4,071 training images, 872 validation images, and 874 testing images, resulting in 20,355, 4,360, and 4,370 images for training, validation, and testing, respectively. The training set was used to optimize the parameters of TriFusionNet, whereas the validation set was employed to monitor model performance, guide the learning process, and determine the optimal model checkpoint. The independent test set was completely excluded from model training and model selection and was used only for the final performance evaluation. This balanced and systematically partitioned dataset provides a consistent experimental setting for assessing the classification performance and generalization capability of the proposed TriFusionNet framework.

C. Image Preprocessing and Data Augmentation

All input images underwent the same pre-processing pipeline before being fed into the proposed TriFusionNet architecture to make them compatible with the pre-trained EfficientNet-B0 and ResNet18 networks. All the RGB images were resized to 224 × 224 pixels. These images were then normalization with the mean and standard deviation of the ImageNet database. This normalization enables a uniform input distribution and allows the optimization of the pretrained feature-extraction networks to be stable. Data augmentation was used exclusively for the training set to enhance the generalization ability of the proposed model and to reduce the chance of overfitting. Random horizontal flipping, random rotation, scaling and colour jittering were all performed in the augmentation pipeline. These transformations introduce variations in the visual samples (Orientation, scale, and Appearance) while maintaining the semantic identity of the waste category, thereby enhancing the visual diversity of the training samples. No data augmentation was used for validation and test sets. The validation and test images were resized and normalized, but this did not affect the way they were evaluated, as they remained representative of unseen data. All the images were preprocessed and then passed simultaneously to the EfficientNet-B0, ResNet18, and Custom CNN branches.

D. TriFusionNet Architecture

The EfficientNet-B0 branch removes its original classifier and provides a 1280-dimensional feature vector. The ResNet18 branch removes its original fully connected classifier and provides a 512-dimensional representation. Three convolutional blocks with batch normalization and ReLU activation, pooling, and adaptive average pooling are included in the custom CNN. producing a 128-dimensional feature representation.

The fused feature representation is:

$$F_{fusion} = [F_E \parallel F_R \parallel F_C] \in \mathbb{R}^{1280+512+128} = \mathbb{R}^{1920}.$$

where $F_E \in \mathbb{R}^{1280}$, $F_R \in \mathbb{R}^{512}$, and $F_C \in \mathbb{R}^{128}$. The classification head consists of the layers 1920 → 512 → 256 → 5, with batch normalization, ReLU activation, and dropout as implemented in the saved model.

Table 1. Architecture of the Proposed TriFusionNet Model

Component	Input	Output	Description
EfficientNet-B0	224 × 224 × 3	1280	High-level feature extractor
ResNet18	224 × 224 × 3	512	Complementary residual feature extractor
Custom CNN	224 × 224 × 3	128	Task-specific local feature extractor
Feature Fusion	–	1920	Heterogeneous feature concatenation

FC1	1920	512	Dense + BatchNorm + ReLU + Dropout
FC2	512	256	Dense + ReLU + Dropout
Output	256	5	Five-class classification

E. Proposed TriFusionNet Architecture

The hybrid feature extraction mechanism is the main part of the proposed model and comprises three parallel branches:

EfficientNet-B0 Branch: EfficientNet-B0 serves as the primary semantic feature extractor in the proposed framework. It employs a compound scaling strategy that jointly optimizes the network's depth, width, and input resolution, enabling high classification performance with reduced computational complexity. The final classification layer of the pre-trained EfficientNet-B0 model is removed, and the resulting 1280-dimensional feature vector is utilized for subsequent feature fusion and classification.

ResNet18 Branch: ResNet18 uses identity shortcut connections to learn complementary residual features, propagate gradients, and learn deep features effectively. A fully connected layer is replaced by an identity mapping, resulting in a 512-dimensional representation of features.

Custom CNN Branch: The third arm of the TriFusionNet is a lightweight Custom CNN to retrieve local visual representations for specific tasks from a mass of images. The network is comprised of 3 convolutional blocks. There is a within each block. The convolutional layer is followed by a batch normalization layer and the ReLU layer. The first two blocks are followed by two max-pooling layers with a stride of 2. The first 2 blocks are followed by 2 max-pooling layers with a stride of 2 to reduce the spatial resolution progressively. The number of convolutional filters rises from32 to 64 and then to 128, allowing the network to reach the following stage: learn more complex feature representations. The first convolutional layer has a filter size of $3 \times 3 \times 3$. The filter size of the first convolutional layer is $3 \times 3 \times 3$ for an input image of size $224 \times 224 \times 3$. The 32 filters in the layer are all 3×3 filters with padding 1of that generate a layer. $224 \times 224 \times 32$ representation. This can be reduced by using max pooling. $112 \times 112 \times 32$. The second convolutional layer has 64 filters. creates a representation of size $112 \times 112 \times 64$ and then applies max pooling to get rid of 64 dimensions. obtain $56 \times 56 \times 64$. The third convolution uses 128 filters; this gives a representation of $56 \times 56 \times 128$. Batch normalization and ReLU activation are applied after each convolution. Finally, adaptive average pooling (AAP) decreases the spatial dimensions to 1×1, and the flattened representation is. 128-dimensional feature vector F_C.. he detailed architecture of the Custom CNN is presented in Table 2.

Table 2. Architecture of the Custom CNN branch.

Layer	Configuration	Kernel/Pool	Output Size
Input	RGB image	–	$224 \times 224 \times 3$
Conv2D-1	32 filters	3×3	$224 \times 224 \times 32$
BatchNorm + ReLU	–	–	$224 \times 224 \times 32$
MaxPool-1	–	2×2	$112 \times 112 \times 32$
Conv2D-2	64 filters	3×3	$112 \times 112 \times 64$
BatchNorm + ReLU	–	–	$112 \times 112 \times 64$
MaxPool-2	–	2×2	$56 \times 56 \times 64$
Conv2D-3	128 filters	3×3	$56 \times 56 \times 128$
BatchNorm + ReLU	–	–	$56 \times 56 \times 128$
Adaptive Avg. Pool	–	1×1	$1 \times 1 \times 128$
Flatten	–	–	128

The resulting Custom CNN feature vector is combined with the feature representations extracted by EfficientNet-B0 and ResNet-18. Specifically, EfficientNet-B0 produces a 1280-dimensional feature vector F_E, ResNet-18 produces a 512-dimensional feature vector F_R, and the Custom CNN produces a 128-dimensional feature vector F_C. The three representations are concatenated as

$$F_{fusion} = [F_E \| F_R \| F_C] \in \mathbb{R}^{1920}.$$

Thus, the fused representation contains $1280 + 512 + 128 = 1920$ features and is subsequently passed to the classification head.

Feature Fusion: The features obtained with EfficientNet, ResNet, and the custom CNN are concatenated using a concatenation operation. This hybridization step forms a composite feature vector that has both low-level and high-level information. The model combines properties of several architectures to obtain a more discriminative and richer representation, which enhances its performance in classification.

$$F_{fusion} = \mathrm{Concat}(F_E, F_R, F_C),$$

Classification Layer: The completely linked (dense) feature vector receives the combined feature vector. layers are finally classified. To avoid overfitting, dropout is used, while label smoothing improves generalization by reducing confidence in specific classes. A Softmax activation is used in the final output layer to produce multi-class classification probabilities.

Algorithm 1: Training and Evaluation of the Proposed TriFusionNet

Input: Training dataset D_{train}, validation dataset D_{val}, test dataset D_{test}
Output: Optimized hybrid model M and evaluation metrics

Load D_{train}, D_{val} and D_{test};
Resize all images to 224×224;
Apply data augmentation only to training images;
Normalize all images using ImageNet statistics;
Initialize pretrained EfficientNet-B0 and remove its classifier;
Initialize pretrained ResNet18 and remove its fully connected layer;
Initialize the Custom CNN feature extractor;
Freeze EfficientNet-B0 and ResNet18 parameters during the initial training stage;
Extract feature dimensions: $F_E \in \mathbb{R}^{1280}$, $F_R \in \mathbb{R}^{512}$, and $F_C \in \mathbb{R}^{128}$;
Construct the feature-fusion representation: $F_{fusion} = [F_E \parallel F_R \parallel F_C]$;
The fused feature dimension is $1280 + 512 + 128 = 1920$;
Construct the classification head $(1920 \rightarrow 512 \rightarrow 256 \rightarrow 5)$;
Apply Batch Normalization, ReLU activation and Dropout according to the implemented architecture;
Initialize CrossEntropy Loss with label smoothing (0.05);
Initialize AdamW optimizer with initial learning rate ($lr = 10^{-3}$);
Initialize ReduceLROnPlateau scheduler;
Initialize Early Stopping with patience = 7;
Set $BestValidationAccuracy \leftarrow 0$;
for $epoch = 1$ **to** E **do**
  if $epoch = 6$ **then**
    Unfreeze EfficientNet-B0 and ResNet18;
    Update optimizer learning rate to 10^{-4};
  Set model to training mode;
  foreach *mini-batch* (X, Y) *in* D_{train} **do**
    Extract feature vector F_E using EfficientNet-B0;
    Extract feature vector F_R using ResNet18;
    Extract feature vector F_C using the Custom CNN;
    Fuse the extracted features: $F_{fusion} = [F_E \parallel F_R \parallel F_C]$;
    Pass F_{fusion} through the classification head;
    Predict output probabilities for the five waste classes;
    Compute the label-smoothed CrossEntropy loss;
    Perform backpropagation;
    Apply gradient clipping;
    Update model parameters using AdamW;
  Evaluate the model using D_{val};
  Compute validation loss and validation accuracy;
  Update the ReduceLROnPlateau learning-rate scheduler;
  if *Validation Accuracy* > *BestValidationAccuracy* **then**
    Save the current model weights;
    Update *BestValidationAccuracy*;
    Reset the early-stopping counter;
  else
    Increment the early-stopping counter;
  if *Early-stopping counter* ≥ 7 **then**
    Break training;
Load the model weights corresponding to the best validation performance;
Evaluate the optimized model on the independent test dataset D_{test};
Compute Accuracy, Precision, Recall, F1-score and Confusion Matrix;
Compute ROC-AUC using the test-set prediction probabilities;
Generate training and validation accuracy/loss curves;
Generate Grad-CAM visualizations for model interpretability;
Save the optimized model and evaluation results;
return Optimized TriFusionNet model M;

1

F. Training and Evaluation Algorithm

Algorithm 1

First, the training, validation and test sets are loaded. Every input picture is scaled to 224 by 224 pixels, and data augmentation is performed only on the training set. Images are normalized with values of mean/standard deviation from ImageNet. The three branches then extract a feature vector from each of the input images: EfficientNet-B0 extracts a 1280-dimensional feature vector F_E, ResNet-18 extracts a 512-dimensional feature vector F_R, and the Custom CNN extracts a 128-dimensional feature vector F_C. These representations are merged to create a single feature vector of length 1920:

$$F_{fusion} = [F_E \parallel F_R \parallel F_C] \in \mathbb{R}^{1920},$$

where

$$1280 + 512 + 128 = 1920.$$

The fused representation is fed into the classification head of the network, having the dimensions of 1920 → 512 → 256 → 5. The pretrained EfficientNet-B0 and ResNet-18 parameters are fixed throughout the first training phase. and the parameters of the Custom CNN and classification layers are learned. The unfrozen branches in the main network are unfrozen and started to be fine-tuned at epoch 6, while learning rates are lowered from 10^{-3} to 10^{-4}. The model is trained with AdamW, the learning rate scheduler ReduceLROnPlateau and gradient clipping with label-smoothed cross-entropy loss. A patience of 7 epochs with an early stopping criterion based on validation accuracy is used. The model parameters that give maximum validation accuracy are kept and then tested with accuracy, precision, recall, F1-score, confusion matrix and ROC-AUC on the independent test set. A training curve as well as a validation curve is also produced, and Grad-CAM is employed to visualize the region of the image that is relevant to the model's predictions.

4. Experimental Setup

A. Implementation and Training Details

It is proposed to implement the TriFusionNet using the PyTorch deep learning framework and train the model on a GPU system. The images were resized to 224×224 and pre-processed with ImageNet mean and std. We used random horizontal flip, rotation, scaling and color jittering only on the training set to enhance the generalization ability of the model. AdamW optimizer was used with a learning rate of 1×10^{-3} in the training of the network. The batch size is 32, and it is used for each batch. The objective function used was Cross-Entropy Loss with label smoothing (0.05). All layers of EfficientNet-B0 and ResNet18 were unfrozen, and the learning rate was set to 1×10^{-4} during the fine-tuning stage for optimizing the entire end-to-end flow. This learning rate was then adjusted using the ReduceLROnPlateau scheduler with respect to the validation loss. To promote the stability of optimization, gradient clipping was adopted with a clip value of 0.3. To avoid overfitting, an Early Stopping method with a patience value of 7 epochs was implemented. The best model based on the highest validation accuracy was saved and tested using the independent test data.

B. Hyperparameter Configuration

Table: 3. Training and Hyperparameter Configuration of TriFusionNet

Parameter	Value
Input image size	224×224
Batch size	32
Optimizer	AdamW
Initial learning rate	1×10^{-3}
Fine-tuning learning rate	1×10^{-4}
Weight decay	1×10^{-4}
Loss function	Cross-Entropy + label smoothing (0.05)
Dropout	0.3
Number of classes	5
Normalization	ImageNet mean/std
Data augmentation	Flip, rotation, scaling, color jitter
Early stopping	Patience = 7
Learning-rate scheduler	ReduceLROnPlateau

C. Evaluation Metrics

The performance of the proposed TriFusionNet was evaluated on the independent test set using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC. These metrics were selected to assess both the overall classification performance and the class-wise discriminative capability of the proposed model. Accuracy measures the proportion of correctly classified samples among all test samples and is defined as

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}.$$

Precision measures the proportion of correctly predicted samples among all samples predicted as a particular class:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}.$$

Recall measures the proportion of actual samples correctly identified by the model:

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}.$$

The F1-score is the harmonic mean of precision and recall:

$$\mathrm{F}_1 = 2\frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}.$$

A confusion matrix was used to examine the class-wise distribution of correct and incorrect predictions. In addition, ROC curves and the corresponding Area Under the Curve (AUC) were computed using the predicted class probabilities to evaluate the discriminative capability of the model across the five waste categories. For multi-class evaluation, macro-average and weighted-average precision, recall, and F1-score were also considered. The macro-average assigns equal importance to each class, whereas the weighted-average accounts for the number of samples in each class.

5. Results and Discussion

A. Accuracy and Loss Curve Analysis

Figure 3 shows the accuracy of training and validation using TriFusionNet model for 30 epochs. Both curves show a consistent rising pattern, revealing the performance of the hybrid architecture in effectively learning during training. In the first epochs (1-5), the accuracy of the training set rose rapidly from about 73.5% to 86%, and the accuracy of the validation set rose from about 81.3% to 91.8%, showing that the model acquired discriminative visual features from waste images quite fast. The validation accuracy was above 95%, indicating that the model is well-generalized with efficient feature extraction. During epochs 6 to 10, both curves showed a significant increase. The training loss and validation loss curves of the proposed TriFusionNet model are shown in Figure 3. The loss curves for both are decreasing uniformly during training, which indicates the proper optimization of the network parameters. The training loss curve dropped rapidly from around 0.83 to 0.57 in the first five epochs, and so did the validation loss curve from around 0.65 to 0.43, both during the early epochs of training, showing rapid convergence in early epochs. This significant decrease shows that the model was able to learn meaningful feature representations and to achieve a significant reduction in classification errors.

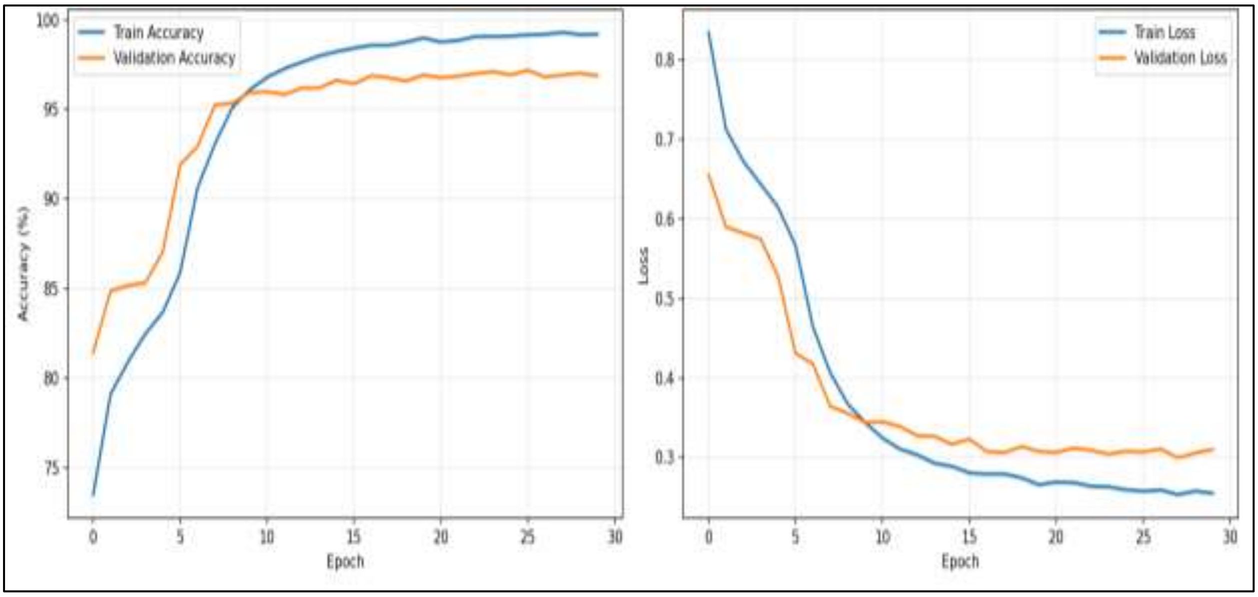

Fig.3. Training and validation accuracy curves.

The training and validation curves are closely aligned with minimal gap, indicating a relatively small gap between training and validation performance, and suggesting limited overfitting under the evaluated training conditions.

B. Confusion Matrix

The proposed TriFusionNet model classification accuracy on the 5 waste categories (glass, metal, organic, paper, plastic) is shown in the confusion matrix in Figure 3. The model was able to obtain an overall test accuracy of 97.16%, which shows its efficiency in discriminating between the categories of recyclable and organic wastes. A high concentration of values along the main diagonal suggests that most of the waste samples have been classified correctly, with a small proportion of misclassified samples. This demonstrates how well the suggested feature-fusion approach learns discriminative representations of different waste items. The organic class showed the best classification performance among all the classes. Of 874 test images, 858 images were correctly classified, with only a few test images being misclassified as paper (6), plastic (3), metal (4), and glass (3). The recognition ability of the glass category was also found to be good, with 850 samples recognized. There were a few glass samples misidentified as plastic (8), paper (6), organic (4) and metal (8). Likewise, the metal class made a total of 845 correct predictions from 874 test images. Most of the misclassifications were directed towards plastic (8), organic (5), paper (6), and glass (5). Although there were some minor errors, the class performed well across the board, with precision and recall values near 0.97, and the model was able to correctly identify metallic objects even when they shared reflective characteristics with other recyclable materials. In the paper category, the model was able to correctly classify 845 samples. The major misclassifications were in the areas of plastic (10); organic (6); metal (9); and glass (7). These errors may be attributed to the similarities between paper-based products and light plastic packaging, particularly when there is a folded or crumpled object or part of an object in the pictures. However, the precision of 0.97, recall of 0.96 and F1 score of 0.97 were quite good throughout the class. For Plastic, the model correctly identified 848 images; more plastic images were misclassified as metal (7), paper (7), organic (4), and glass (8). In general, the confusion matrix shows that most of the errors in the classification are between materials that are highly alike in terms of their visual characteristics, not between materials and organic waste. This implies that the proposed hybrid architecture has strong class separability and can capture high-level semantic information. Once more, the weighted average and macro average of accuracy, recall, and F1-scores are 0.97, indicating that the model exhibits no discernible bias toward any class. TriFusionNet utilizes EfficientNet-B0, ResNet18 and the Custom CNN to take advantage of complementary feature representations. EfficientNet-B0 is responsible for efficient global feature extraction, ResNet18 is for boosting residual feature learning, and the Custom CNN is for gathering more local texture information. Their fused representation also enhances discrimination among the hard-to-separate classes, resulting in high classification accuracy and fair performance on the entire test image set of 4370 images.

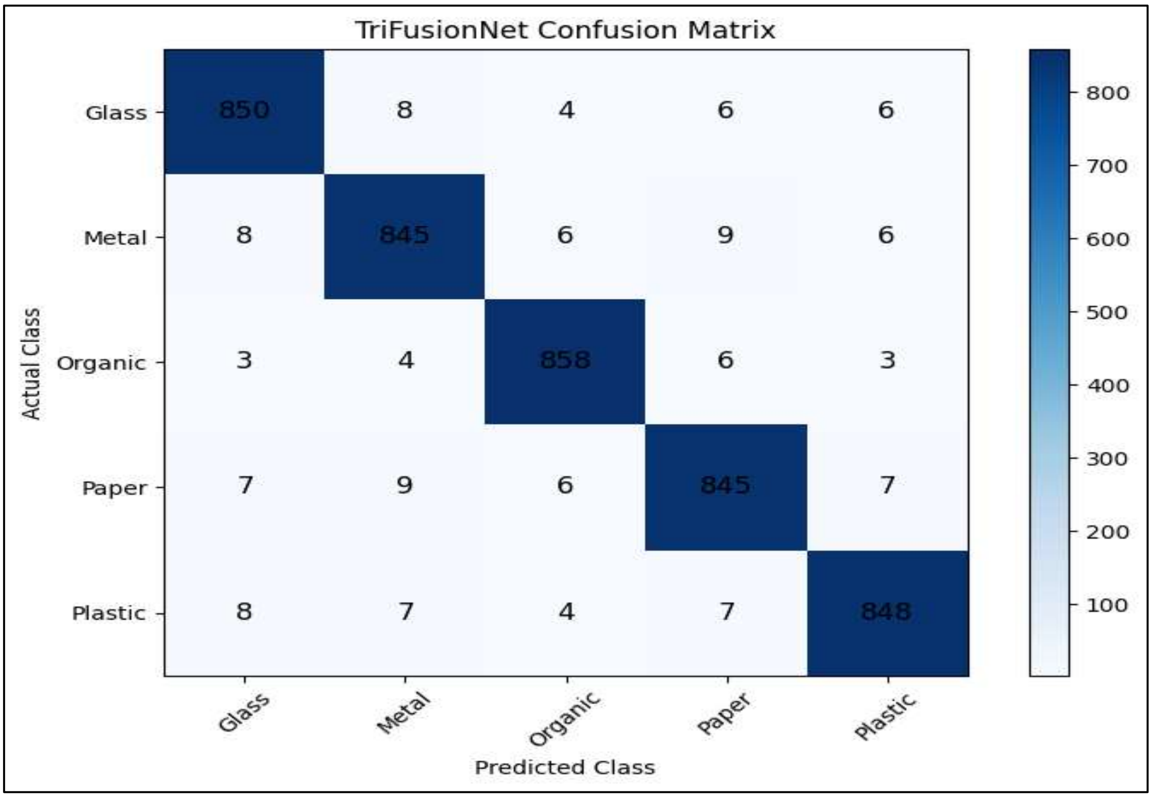

Fig. 4. Confusion matrix

C. Classification Report Analysis

Table 4: To analyze the classification ability of the proposed TriFusionNet model, precision, recall, F1-measure, and overall accuracy were employed as classification performance metrics. These metrics can give a thorough evaluation of the accuracy of the model in accurately categorizing every type of garbage while simultaneously reducing false positive and false negative predictions. The evaluation was done on the test set containing 4,370 images divided into five waste classes: glass, metal, organic, paper, and plastic (874 images per class). The overall classification results show that the proposed model successfully obtained an overall accuracy of 97% and gave good generalization on the unseen test samples. Furthermore, the model demonstrated consistent performance across all waste categories without any bias toward particular waste categories, as seen by the macro-average and weighted-average accuracy, recall, and F1-score values of 0.97. The best performance was obtained in the class of organic waste (precision 0.97, recall 0.99 and F1 0.98). The remarkable recall score shows how successful the suggested model is in correctly identifying almost all the samples from the organic waste with minimal false negatives. The unique visual property of the organic waste could be the reason for this superior performance, as the hybrid feature extraction framework could be able to identify its patterns of discrimination. Additionally, the glass class had a good classification skill with an F1 score of 0.97, recall of 0.97, and accuracy of 0.98. The accuracy of the high accuracy shows that most of the samples that were classified as glass were correctly identified, showing that the model was able to distinguish visually similar recyclable materials from glass. The balanced performance of the metal class was also very good (precision of 0.96, recall of 0.97 and F1-score of 0.97). These similar precision and recall scores indicate

that the proposed network was able to successfully capture the unique features based on the structure and textures of metallic waste, as well as keeping the misclassification rate low. For the paper category, the proposed model obtained a precision of 0.97, recall of 0.96, and F1-score of 0.97. The performance of the recall is slightly lower than that of the organic class, but the overall performance is still very good, suggesting that the feature fusion approach achieves good discrimination against paper waste despite the fact that there are diverse textures, colors, and surface appearances. Out of all the classes, the plastic class performed the worst, with only an accuracy of 0.96, recall of 0.95, and F1-score of 0.95. These values, although the lowest for the class, compare very favorably with the classification capability of other classes. The limited recall may be due to visual similarity between plastic and paper or metal, which leads to a relatively low number of misclassifications. However, the proposed hybrid architecture managed to keep the recognition performance of plastic waste high. The precision and recall values are very close for all five classes, showing that the proposed model, TriFusionNet, does not exhibit bias in the prediction of any of the classes. Furthermore, the identical macro and weighted average values demonstrate that the balanced dataset has helped with uniform classification performance across all classes. Based on the overall experimental results, it can be concluded that the proposed TriFusionNet architecture successfully merges complementary feature representations learned from the pre-trained EfficientNet-B0, ResNet18, and the custom CNN. This hybrid feature fusion mechanism helps the model to achieve better discriminative capability, leading to highly accurate and reliable multi-class waste classification. The outcomes demonstrate how well the suggested approach for putting in place an intelligent waste management system works. where precise identification of recyclable and organic waste becomes crucial for automated waste sorting and environmentally conscious practices.

Table: 4. Class-wise performance of TriFusionNet.

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Glass	97.03	97.25	97.14	874
Metal	96.79	96.68	96.74	874
Organic	97.72	98.17	97.95	874
Paper	96.79	96.68	96.74	874
Plastic	97.47	97.03	97.25	874
Accuracy			**97.16**	**4370**
Macro Avg.	97.16	97.16	97.16	4370
Weighted Avg.	97.16	97.16	97.16	4370

D. Class-wise analysis

The classification outcomes of the suggested model, TriFusionNet, evaluated per class, show consistently good classification accuracy across all five waste classes. The glass class had a Precision of 0.98, Recall of 0.97 and F1-Score of 0.97, indicating high precision of identification with a low rate of false positive predictions. Similarly, the metal class had a precision of 0.96, a recall of 0.97, and an F1 score of 0.97, showing that recognition of metallic waste is balanced and reliable. The model's overall performance was excellent, with the organic class having a precision of 0.97, a recall of 0.98 and the highest F1 score of 0.98 of all the classes, with very few samples being missed. The model achieved a high precision of 0.97, a recall of 0.96 and an F1 score of 0.97 for the paper category, demonstrating its ability to accurately identify paper waste even when there are differences in paper texture and appearance. The precision, recall, and F1-scores of the plastic class were the lowest among the five classes at 0.96, 0.95, and 0.95, respectively. The slightly lower performance might be explained by the visual variety of plastic waste and the ease of contamination with other recyclables. However, the high precision, recall, and F1-score values in all categories show that the proposed TriFusionNet is effective in learning the discriminative feature representations and offers accurate, balanced, and reliable multi-class waste classification.

E. Analysis of Misclassified Categories

The proposed TriFusionNet model has high classification performance in all five categories of waste, which means only a small number of samples are misclassified. Very few errors occurred when considering the organic class (0.98), suggesting that very few samples of organic waste were misclassified in any other class. The other classes, glass, metal, and paper, also performed well with F1-scores of 0.97, which means that there were few misclassified instances. The model performed worst in the plastic category, with a low recall (0.95) and F1-score (0.95) among all categories, indicating that plastic waste was the most challenging category for the model. This lower performance could be explained by the number and diversity of plastic objects that can appear, feel, be coloured and be shaped in a similar way to other materials that are also recyclable. In general, high precision and recall rates for each class imply that the number of misclassified samples was relatively small, indicating the robustness and reliability of the proposed framework of TriFusionNet for multi-class waste classification.

1. Macro and Weighted Average Performance

The precision, recall and F1 macro-average in the proposed model were 0.97. The macro-average metric is computed across each class regardless of the number of samples, giving equal weight to each class and thus providing an unbiased measure of the overall classification. The same macro-averages reflect the same predictive power across every waste category with no bias towards any waste category. The balanced behavior shows that the hybrid feature extraction strategy works well on various types of recyclable and organic waste. The weighted-average precision, recall, and F1-score were also 0.97. The weighted average takes the number of samples in each class into consideration when calculating the overall performance, unlike the macro average. Both the macro-average and the weighted-average metrics are similar since all the waste categories have the same number of test images (874 samples). This agreement validates that the proposed model achieves stable and uniform classification performance within the balanced dataset, while eliminating the bias towards any waste category.

2. Implications for Smart Waste Management

The proposed model's outstanding classification performance proves its potential applications to intelligent waste management systems. Accurate waste classification allows for more efficient automated sorting, reduces reliance on manual sorting and improves the efficiency of recycling operations. Furthermore, the uniformity of results for various waste categories suggests that the system is suitable for a wide range of waste management applications, including smart waste bins, recycling centres, and waste management monitoring systems. Through its use, it can help to ensure sustainable waste management, including waste segregation, optimizing the recovery of resources, enabling circular economy practices, and contributing to sustainable environmental development in the long term.

F. ROC-AUC Analysis

The data in Figure 5 are the receiver operating characteristic (ROC) curves for the one-vs-rest classification. The five waste categories have characteristic (ROC) curves. The proposed TriFusionNet exhibited good discriminative ability for all The AUC values of the classes were glass (0.996), metal (0.996), and plastic (0.996). 0.999 for organic, 0.997 for paper. The micro-average and macro-average AUC were 0.996 and 0.997, respectively, respectively. The ROC curves stay near the top left corner, suggesting a high true positive rate and low false positive rate. Many decision thresholds across.

Of the five categories, the highest was obtained for the category of organic waste. The highest-class separability of 0.999 was achieved with (0.999). In contrast, the lowest (0.995) was achieved with plastic waste, but the value was still relatively good. A score of >0.80 is a strong discriminative score. This observation is in line with the results of class classification; plastic also had the lowest recall and F1-score because it had larger visual similarities to other waste materials. Overall, the ROC-AUC analysis, as demonstrated by the following data, TriFusionNet learns very well. Five-class waste classification discriminative representation task.

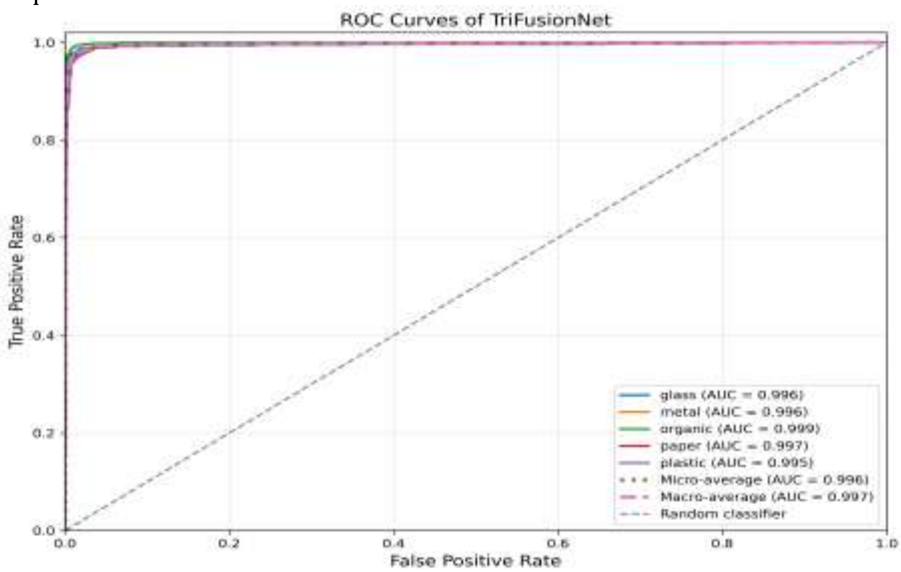

Fig. 5: One-vs-rest ROC curves of TriFusionNet for the five waste categories. The class-wise AUC values are 0.996 for glass, 0.996 for metal, 0.999 for organic, 0.997 for paper, and 0.995 for plastic. The micro-average and macro-average AUC values are 0.996 and 0.997, respectively.

G. Grad-CAM Visualization

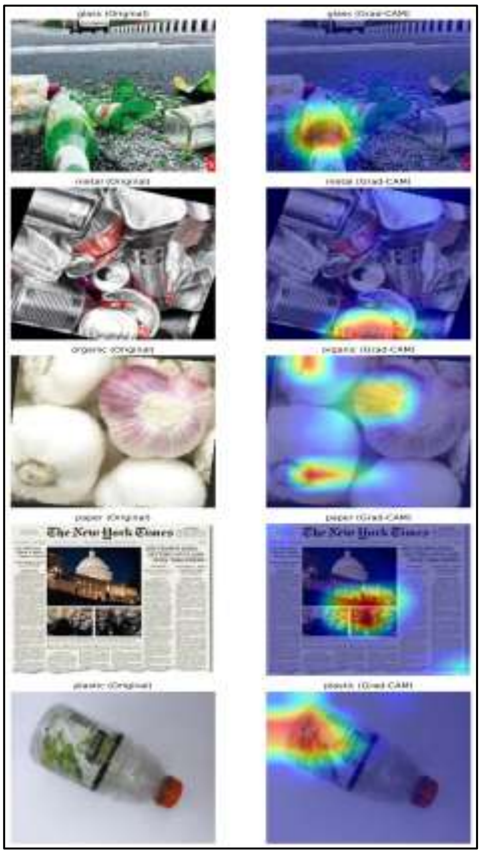

Fig.6. Grad-CAM visualization showing important regions influencing predictions.

This figure 6. To make the proposed TriFusionNet model more interpretable, Gradient-weighted Class Activation Mapping (Grad-CAM) was used to visualize the regions of this image that were most important for making the classification judgments. The images in the left column are the original waste images, and the images in the right column are the Grad-CAM heat maps corresponding to the images in the left column, as shown in Figure 5. The heatmaps show red and yellow areas when the feature is the most influential and green and blue areas when the feature has a relatively low influence. The activation maps for all five waste categories focus on the target waste objects and exclude background information, suggesting that meaningful visual patterns have been learned and are useful for the classification task. The localization results clearly show that the hybrid feature extraction framework captures discriminative semantic features essential for recognizing waste. Also, a class-wise analysis confirms the reliability of the proposed model. In the glass class, the activation is focused on the body of the bottle, with emphasis on the shape and the texture. In the metal category, the model focuses mainly on the metallic can, emphasizing reflective surfaces and structural characteristics. When comparing the model performance with the organic sample, strong activation is observed over the garlic bulbs, showing that the model is able to capture the natural texture and shape information well. The greatest attention is seen for the paper class in the central area of the newspaper, where there is printed text and images, and for the plastic class, around the bottle label and upper body area, indicating that these areas offer the most informative visual cues for classification. In summary, the activation maps generally concentrate on regions corresponding to the target waste objects, although some background activation may also be present.

H. Comparison with Baseline Models

Three baseline deep learning models were used to analyze the performance of the suggested TriFusionNet architecture in order to assess its efficacy: the custom CNN, ResNet-18, and EfficientNet-B0. All models were trained and evaluated using the same dataset partition, preprocessing strategy, and evaluation protocol, and their performance was assessed using accuracy, precision, recall, and F1-score. The comparative results are presented in Table 5.

Among the individual baseline models, EfficientNet-B0 obtained the best performance was obtained by the model with an accuracy of 96.80%, 96.82% accuracy in precision, 96.80% accuracy in recall, and 96.80% accuracy in F1-score. The ResNet-18 model's accuracy is 95.24%, and the custom CNN model's accuracy is 79.29%. The relatively lower performance of the custom CNN indicates the advantage of pretrained deep feature representations for the considered five-class waste classification problem.

TriFusionNet showed the best overall performance among the evaluated models. It obtained 97.16% for accuracy, precision, recall, and F1-score. Compared with the strongest individual baseline, EfficientNet-B0,

TriFusionNet improved the classification accuracy by 0.36 percentage points. This improvement indicates that combining complementary feature representations from heterogeneous backbone networks can provide additional discriminative information compared with relying on a single architecture.
In the proposed framework, the lightweight custom CNN is intended to capture task-specific local visual patterns, EfficientNet-B0 provides efficient high-level semantic representations, and ResNet-18 contributes residual feature representations. These heterogeneous feature vectors are subsequently integrated through the proposed fusion mechanism to form a joint representation for the five waste categories, namely glass, metal, organic, paper, and plastic. The results therefore demonstrate the potential of multi-backbone feature integration for improving recyclable-waste image classification.

Table.5. Performance comparison of the baseline models and the proposed TriFusionNet.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Custom CNN	79.29	79.28	79.29	79.25
EfficientNet-B0	96.80	96.82	96.80	96.80
ResNet-18	95.24	95.26	95.24	95.24
TriFusionNet	**97.16**	**97.16**	**97.16**	**97.16**

6. Ablation Study

A series of ablation experiments was performed on seven different setups to understand the individual and joint contribution of Custom CNN (A), EfficientNet-B0 (B) and ResNet-18 (C). The individual branches did best with EfficientNet-B0, which had a 96.80% accuracy, followed by ResNet-18 at 95.24% and the Custom CNN at 79.29%. There was no consistent improvement in performance in the pairwise configurations when compared with the individual backbones. The A+B and A+C configuration examples got accuracies of 96.61% and 94.74%, respectively, and the B+C configuration got 96.70% accuracy.
The highest performance is seen in the full 3-branch configuration, A+B+C (TriFusionNet), The proposed model achieved an improved accuracy, precision, recall and F1-score of 97.16%, which is 0.36% higher than the highest single model of EfficientNet-B0, and 0.46 points better than the best pairwise configuration (B+C). The overall performance of the pairwise combinations was not consistent, but the three-branch configuration was the most effective. The outcomes demonstrate that the fused and diverse feature representations employed in the TriFusionNet complement each other.

Table.6 Ablation study of the proposed TriFusionNet framework.

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Custom CNN (A)	79.29	79.28	79.29	79.25
EfficientNet-B0 (B)	96.80	96.82	96.80	96.80
ResNet-18 (C)	95.24	95.26	95.24	95.24
A+B	96.61	96.63	96.61	96.62
A+C	94.74	94.79	94.74	94.74
B+C	96.70	96.72	96.70	96.71
A+B+C (TriFusionNet)	**97.16**	**97.16**	**97.16**	**97.16**

7. Comparison with published Approaches

In this work, we proposed the TriFusionNet framework with representative deep-learning approaches reported in the literature, which are presented in Table 7. The chosen studies include CNN, hybrid networks, deep-feature fusion, heterogeneous feature integration, and ensemble-learning approaches. This comparison gives a general idea of the framework that is being proposed and shows the progression away from a single backbone to multiple sources of learned representations. The first of the aforementioned fusion-based methods was Ahmad et al. [19], which fused deep features and score-level information with a double fusion method and achieved a performance improvement of 3.58% over the state-of-the-art method. They also show that it is possible to improve waste-classification performance by using complementary information from multiple deep-learning models. In the same way, Shi et al. [20] developed a multi-layer hybrid convolutional neural network and achieved an accuracy of 92.60% on the TrashNet dataset, which indicated the potential of hybrid architectures by considering the complexity of the model. Waste classification was further enhanced by Mao et al. [21] by employing an optimized DenseNet121 network structure, data augmentation and optimization using a genetic algorithm (GA), giving an accuracy of 99.60%. The studies suggest that architectural optimization, feature integration and transfer learning can significantly impact the classification performance.
More like the proposed framework, ECCDN-Net fused DenseNet201 and ResNet18 feature representations and attained a 96.10% accuracy for a two-class dataset for organic vs recyclable waste. This outcome serves as proof of the ability of heterogeneous CNN features to be used to complement the information used for the classification of waste. The present study, however, focuses on 5 categories of recyclable waste, while only

two categories were considered by the ECCDN-Net study. Recently, ensemble learning work by Alkılınç et al. [22] also showed the superior performance of averaging and weighted averaging of various CNN predictions. All of the above studies argue in favor of the use of multi-model learning and also suggest that it is not always the most efficient learning approach when using fusion, as it will depend on the chosen architectures, fusion approach, the data set properties, and the evaluation protocol.

The proposed TriFusionNet was able to achieve 97.16%, 97.16%, and 97.16% for accuracy, precision, and recall, respectively and macro F1 was 97.16% on the five-class dataset in the present study. The framework is a lightweight custom CNN with EfficientNet-B0 and ResNet-18 to learn complementary feature representations. The proposed architecture is based on multiple representations derived from different architectures with different structural properties, as compared to a single backbone approach. The results thus show competitive performance of the proposed multi-backbone approach for classification of 5-class recyclable waste.

In the experimental setting of the present study, the accuracy, precision, recall and macro F1-score of the TriFusionNet on the five-class unified dataset were 97.16%. This is better than EfficientNet-B0 (96.80%) and ResNet18 (95.24%) individual baseline results, with the same experimental setting. But it should not be concluded that TriFusionNet is always superior to published methods reported in Table 7 since they are conducted on different datasets and experimental setups. Specifically, Mao et al. [21] claimed high accuracy of 99.60% on TrashNet, but the results were achieved by adopting a different dataset (with six classes) and an optimized experimental framework. Thus, the comparison is mainly showing the competitive performance in the five-class unified dataset investigated in the present study.

Table.7 Comparison of the proposed TriFusionNet with representative published deep-learning approaches for waste classification.

Study	**Method / Approach**	**Dataset**	**Classes**	**Reported accuracy (%)**	**Other reported metric / finding**	**Relevance to present study**
Ahmad et al. [19]	Double fusion of deep features and score-level information	Waste image datasets	—	**NR**	3.58% improvement over previous state of the art	Demonstrates the potential of combining complementary deep representations
Shi et al. [20]	Multilayer Hybrid CNN (MLH-CNN)	TrashNet	6	**92.6**	—	Shows the effectiveness of hybrid CNN architectures
Mao et al. [21]	Optimized DenseNet121 with data augmentation and GA-based optimization	TrashNet	6	**99.6**	Grad-CAM-based interpretation	Provides a strong transfer-learning-based comparison
ECCDN-Net [23]	DenseNet201 + ResNet18 feature fusion	24,705 waste images	2	**96.1**	Binary classification	Closely related to heterogeneous CNN feature fusion
Alkılınç et al. [22]	Deep ensemble learning using averaging and weighted averaging	TrashNet, TrashBox, Waste Pictures, Garbage Classification	Multiple	**96**	Averaging/weighted ensemble	Demonstrates prediction-level model combination
TriFusion Net (Proposed)	Custom CNN + EfficientNet-B0 + ResNet18 feature fusion	Unified dataset of present study	5	97.16	Feature-level heterogeneous fusion + seven-configuration ablation	Integrates semantic, residual, and local feature representations

[a]Reported accuracy on the TrashNet dataset using the weighted ensemble.

Note: Reported accuracies are obtained from the corresponding studies. Direct numerical comparison should be interpreted cautiously because the studies use different datasets, class definitions, image acquisition conditions, and experimental protocols. Therefore, the reported accuracy values should be interpreted as contextual reference points rather than strictly controlled numerical comparisons. In particular, the 3.58% value reported for Ahmad et al. [19] represents an improvement over a previous state-of-the-art result and is therefore not presented as an accuracy value.

8.Conclusion

In this work, a hybrid deep learning framework, TriFusionNet, was proposed, combining EfficientNet-B0, ResNet18, and a custom CNN, and showing its potential in applications for automatic garbage sorting and recycling. Five waste categories (glass, metal, organic, paper, and plastic) were used to create a balanced dataset that was used to train and evaluate the proposed model, and the dataset was obtained from the TrashNet and Waste Classification Dataset. By fusing feature representations from multiple architectures, TriFusionNet learned robust, discriminative visual features for automated waste classification on the test set; TriFusionNet attained 97.16% accuracy, while precision, recall, and macro F1-score were also measured at 97.16%. The ROC-AUC analysis indicated strong discriminative capability across the five waste categories. Furthermore, the Grad-CAM visualizations showed that the model primarily focused on informative regions of the waste objects. These results demonstrate the potential of combining complementary feature representations from the lightweight Custom CNN, EfficientNet-B0, and ResNet-18 for five-class waste image classification. In general, the results indicate that TriFusionNet provides accurate and interpretable predictions for the considered five-class waste classification task. The framework will be expanded to cover other categories of waste and tested with a variety of real-world environmental conditions, and finally optimized to be deployed on lightweight edge devices to enable the smart waste segregation application in real-time.

Reference

1. I. Voukkali, I. Papamichael, P. Loizia, and A. A. Zorpas, “Urbanization and solid waste production: prospects and challenges,” Environmental Science and Pollution Research, vol. 31, no. 12, pp. 17 678–17 689, 2024.
2. A. Soni, P. K. Das, and P. Kumar, “A review on the municipal solid waste management status, challenges and potential for the future indian cities: A. soni et al.” Environment, Development and Sustainability, vol. 25, no. 12, pp. 13 755–13 803, 2023.
3. C. A. Velis, D. C. Wilson, Y. Gavish, S. M. Grimes, and A. Whiteman, “Socio-economic development drives solid waste management performance in cities: A global analysis using machine learning,” Science of the Total Environment, vol. 872, p. 161913, 2023.
4. J. Aberger, S. Shami, B. Ha¨cker, J. Pestana, K. Khodier, and R. Sarc, “Prototype of ai-powered assistance system for digitalisation of manual waste sorting,” Waste management, vol. 194, pp. 366–378, 2025.
5. N. Manakitsa, G. S. Maraslidis, L. Moysis, and G. F. Fragulis, “A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision,” Technologies, vol. 12, no. 2, p. 15, 2024.
6. M. Obaid, H. Younis et al., “Performance comparison of pretrained deep learning models for landfill waste classification,” 2024.
7. D. V. Yevle and P. S. Mann, “Artificial intelligence-based waste management: A review of classification, techniques, issues, and challenges,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 15, no. 2, p. e70025, 2025.
8. Ankush and S. Mankotia, “Hes-net: An efficient and scalable deep learning framework for waste image classification,” International Journal of Computer Information Systems and Industrial Management Applications, vol. 18, no. 3s, pp. 326–340, 2026. [Online]. Available: *https://doi.org/10.70917/ijcisim-2026-2335*
9. A. Jahanbakhshi, M. Momeny, M. Mahmoudi, and P. Radeva, “Waste management using an automatic sorting system for carrot fruit based on image processing technique and improved deep neural networks,” Energy Reports, vol. 7, pp. 5248–5256, 2021.
10. M. Kaya, S. Ulutu¨rk, Y. C¸ etin Kaya, O. Altıntas¸, and B. Turan, “Optimization of several deep cnn models for waste classification,” Sakarya University Journal of Computer and Information Sciences, vol. 6, no. 2, pp. 91–102, 2023.
11. U. K. Lilhore, S. Simaiya, S. Dalal, M. Radulescu, and D. Balsalobre-Lorente, “Intelligent waste sorting for sustainable environment: A hybrid deep learning and transfer learning model,” Gondwana Research, vol. 146, pp. 252–266, 2025.
12. M. M. Hossen, A. Ashraf, M. Hasan, M. E. Majid, M. Nashbat, S. B. A. Kashem, A. K. A. Kunju, A. Khandakar, S. Mahmud, and M. E. H. Chowdhury, “Gcdn-net: Garbage classifier deep neural network for recyclable urban waste management,” Waste Management, vol. 174, pp. 439–450, 2024.
13. P. Davis, F. Aziz, M. T. Newaz, W. Sher, and L. Simon, “The classification of construction waste material using a deep convolutional neural network,” Automation in Construction, vol. 122, p.

103481, 2021.
14. P. K. Sarswat, R. S. Singh, and S. V. S. H. Pathapati, "Real time electronic-waste classification algorithms using the computer vision based on convolutional neural network (cnn): Enhanced environmental incentives," Resources, Conservation & Recycling, vol. 207, p. 107651, 2024.
15. M. Nahiduzzaman, M. F. Ahamed, M. Naznine, M. J. Karim, H. B. Kibria, M. A. Ayari, A. Khandakar, A. Ashraf, M. Ahsan, and J. Haider, "An automated waste classification system using deep learning techniques: Toward efficient waste recycling and environmental sustainability," Knowledge-Based Systems, vol. 310, p. 113028, 2025.
16. P. Das, Q. Zeng, J.-B. Sirven, J.-M. Lee, and J.-C. P. Gabriel, "Ai-enabled spectral classification of plastic resins from e-waste via laser-induced breakdown spectroscopy for advanced sorting applications," Resources, Conservation & Recycling, vol. 226, p. 108660, 2026.
17. F. Ozkefe, "Trashnet," 2022, kaggle dataset. [Online]. Available: *https://www.kaggle.com/datasets/feyzazkefe/trashnet*
18. S. Sekar, "Waste classification data," 5/18/2024, kaggle dataset, accessed jan 23, 2026. [Online]. Available: https://www.kaggle.com/datasets/techsash/ waste-classification-data
19. K. Ahmad, K. Khan, and A. Al-Fuqaha, "Intelligent fusion of deep features for improved waste classification," IEEE access, vol. 8, pp. 96 495–96 504, 2020.
20. C. Shi, C. Tan, T. Wang, and L. Wang, "A waste classification method based on a multilayer hybrid convolution neural network," Applied Sciences, vol. 11, no. 18, p. 8572, 2021.
21. W.-L. Mao, W.-C. Chen, C.-T. Wang, and Y.-H. Lin, "Recycling waste classification using optimized convolutional neural network," Resources, Conservation and Recycling, vol. 164, p. 105132, 2021.
22. A. Alkılınc¸, F. Y. Okay, ˙I. Ko¨k, and S. O¨ zdemir, "Deep ensemble learning model for waste classification systems," Sustainability, vol. 18, no. 1, p. 24, 2025.
23. M. S. B. Islam, M. S. I. Sumon, M. E. Majid, S. B. A. Kashem, M. Nashbat, A. Ashraf, A. Khandakar, A. K. A. Kunju, M. Hasan-Zia, and M. E. Chowdhury, "Eccdn-net: A deep learning-based technique for efficient organic and recyclable waste classification," Waste Management, vol. 193, pp. 363–375, 2025