



HES-Net: An Efficient and Scalable Deep Learning Framework for Waste Image Classification

Ankush¹, Sunil Mankotia²

¹Department of Computer Science, Himachal Pradesh University, Shimla.

²University College of Business study (UCBS) Avalodge, Shimla.

E-mail(s): ankushkabir6@gmail.com; mankotia.sunil@gmail.com;

Abstract: Automated waste classification has become an essential component of intelligent waste management systems and sustainable environmental monitoring. Although deep convolutional neural networks (CNNs) achieve high classification accuracy, many existing architectures require significant computational resources, limiting their deployment in resource-limited systems like embedded systems, IoT devices and mobile devices. To address this challenge, this study proposes HybridEffiShuffleNet (HES-Net), a lightweight hybrid deep learning framework for efficient multi-class waste image classification. The proposed architecture integrates EfficientNet-B0 and ShuffleNetV2 into a dual-branch feature-extraction framework to capture complementary semantic and structural representations from waste images. An attention-based feature fusion mechanism is introduced to enhance discriminative feature learning and improve classification robustness. The proposed model was evaluated on a five-class waste image dataset comprising glass, metal, organic, paper, and plastic. Extensive preprocessing and data augmentation techniques were applied to improve model generalization. Experimental results demonstrate that HES-Net achieved an overall classification accuracy of 96.34%, outperforming EfficientNet-B0 (95.20%), ShuffleNetV2 (94.05%), ResNet-18 (91.48%), and ShuffleNetV2-S (76.05%) under identical training conditions. Furthermore, the proposed framework exhibited consistent performance across all waste categories while maintaining computational efficiency suitable for real-world deployment. The findings indicate that HES-Net provides an effective and scalable solution for intelligent waste classification and smart recycling applications.

Keywords: Waste Image Classification, Machine Learning, Deep Learning, Convolutional Neural Network, Efficient neural networks, Automated waste sorting

1. Introduction

The growth of urbanization, industrial development, and consumer-oriented lifestyles has significantly increased global waste generation, leading to serious environmental and public health concerns (Eurostat 2026; World Bank 2026). Inefficient waste disposal and poor recycling practices contribute to land pollution, greenhouse gas emissions, and resource depletion. Effective waste segregation is therefore an essential component of sustainable waste management systems and smart recycling infrastructures. Automated waste classification technologies can improve recycling efficiency, reduce manual labor, and support environmentally sustainable waste processing. In recent years, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have achieved remarkable success in computer vision applications such as image classification, object detection, and scene understanding (Alalibo and Nwazor 2023), (Gurcan and Soylu 2025). CNN-based models automatically learn hierarchical feature representations from images without relying on handcrafted feature extraction methods. Due to



their strong feature-learning capabilities, deep learning models have been widely adopted for automated waste image classification systems (Ahmed et al. 2023). Several architectures, including ResNet, EfficientNet, and ShuffleNet, have demonstrated promising classification performance across various waste categorization tasks (Lilhore et al. 2024), (Ahmad et al. 2025), (Ji et al. 2023). Although existing CNN architectures provide high classification accuracy, many deep learning models require substantial computational resources, memory consumption, and inference time, which limit their practical deployment in resource-constrained environments such as mobile devices, embedded systems, and Internet of Things (IoT) platforms (Almtireen et al. 2025; Dawood 2023). Lightweight architectures are low in computational overhead, but may have lower capability of representing features, especially in the case of visually similar waste categories. Thus, it is a challenge to obtain an effective balance between classification accuracy and computational efficiency is an important challenge in intelligent waste management systems. To overcome this, this paper introduces a novel hybrid deep learning framework for multi-class waste image classification called HybridEffiShuffleNet (HES-Net). The proposed architecture consists of two branches of feature extraction: one for EfficientNet-B0 and another for ShuffleNetV2, to exploit complementary semantic and structural information from a waste image. A five-class waste image dataset, which includes glass, metal, organic, paper, and plastic classes, is used to evaluate the proposed model. The experimental results validate the high classification results of HES-Net that outperforms several baseline architectures, while being computationally efficient for real-world applications for smart waste management. This paper is organized as follows: The related work and the background literature are discussed in Section II. The research gap and challenges that are addressed in this study are discussed in Section III. A proposed methodology and model architecture are described in Section IV. In Section V, the techniques used for preparing and pre-processing the datasets are presented. Experimental results and comparative analysis are discussed in Section VI. Finally, Section VII is the conclusion of the paper, and future directions of research are discussed.

2. Research Gap and Challenges

Recent advancements in deep learning have significantly improved the performance of automated waste image classification systems. Convolutional Neural Network (CNN)-based architectures such as ResNet, EfficientNet, EfficientNetV2, and ShuffleNet have demonstrated promising results in waste categorization and image recognition tasks (Van de Schepop 2025; Lin et al. 2023). However, despite achieving high classification accuracy, several challenges remain unresolved in the practical deployment of these models for real-world intelligent waste management systems. One of the major limitations of existing deep learning architectures is their high computational complexity and memory consumption. Deep CNN models often require substantial processing power, storage capacity, and inference time, restricting their deployment in resource-constrained environments such as embedded systems, edge devices, mobile platforms, and IoT-based smart recycling systems (Almtireen et al. 2025), (Ahmad et al. 2025). Although lightweight architectures such as ShuffleNet reduce computational overhead, they may experience reduced classification performance when distinguishing visually similar waste categories. Another important challenge in waste image classification is the accurate discrimination of classes with overlapping visual characteristics, such as paper and plastic waste. Variations in illumination, object orientation, texture similarity, background complexity, and partial occlusion further increase the difficulty of reliable feature extraction and classification. Lightweight models frequently struggle to capture fine-grained semantic information required for robust multi-class waste categorization.

Furthermore, many existing studies primarily focus on classification accuracy while providing limited analysis of computational efficiency, scalability, model size, inference speed, and practical deployment feasibility (Hasan et al. 2024). Comparative evaluations are often conducted under inconsistent experimental conditions, including different datasets, preprocessing methods, and training configurations, making fair performance assessment difficult. Thus, a single and efficient classification system that can achieve both high accuracy and deployment efficiency for intelligent waste management applications is still needed. In this study, a lightweight hybrid deep learning framework HybridEffiShuffleNet (HES-Net) is proposed to solve these problems for efficient multi-class waste image classification. The proposed model a dual-branch structure, the combination of EfficientNet-B0 and ShuffleNetV2. Additionally, a feature fusion mechanism based on attention is applied to boost the discriminative feature learning and ensure the classification performance is robust for various waste categories. The primary contributions of this study are summarized as follows:

1. To efficiently and scalably classify waste images, a novel hybrid deep learning architecture, HES-Net, is proposed by combining EfficientNet-B0 and ShuffleNetV2 in one dual-branch network.

2. An attention-based feature fusion approach is proposed to achieve effective fusion of complementary feature representations and enhance the classification accuracy.
3. A lightweight computing framework is designed that can be deployed in resource-constrained environments like IoT devices or embedded systems.
4. A detailed comparative assessment is performed against some state-of-the-art CNN architectures (ResNet-18, EfficientNet-B0, EfficientNetV2-S and ShuffleNetV2) in the same experimental setup.
5. Experimental results demonstrate that the proposed HES-Net framework achieves superior classification accuracy while maintaining strong computational efficiency for practical smart waste management applications.

3. Related work

We give a thorough overview of the various techniques used in the past to categorize garbage images in this section. According to the research literature, a number of researchers have concentrated their efforts on using contemporary technologies to achieve favourable results in garbage image categorization.

3.1 Residual Networks (ResNet)

Residual Networks (ReNets) introduce the new idea using the skip connections that let the model learn "residual" information. This technique learn the model of deep neural networks. This method showed that this approach helps the models to train better and achieve higher accuracy. Mostly use ResNet-like ResNet18 and ResNet50 for many computer vision tasks(Lin et al. 2023). ResNet18 is a smaller version, but it gives good accuracy. It is often chosen a strong bade line that offers good accuracy, balance between depth, and computational cost. The Residual Networks underlying mapping $H(x)$, to learn a residual function $F(x)$ are defined as

$$F(x) = H(x)c - x$$

Also, they are written as:

$$H(x) = F(x) + x$$

x denotes the input features map, and $H(x)$ represents the output mapping. In Backpropagation and Gradient Stability, the gradient of the loss function L with respect to input x of the residual block is expressed as

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \left(1 + \frac{\partial F(x)}{\partial x} \right)$$

This equation shows that the gradient consists of two components: the directly identified term 1 and the gradient residual function $\frac{\partial F(x)}{\partial x}$. The identity term allows gradients to flow directly through the network without attenuation, even when the residual gradients are small. This property mitigates the vanishing gradient problem and enables the training of deep architectures.

3.2 Compound Scaling and the EfficientNet Family

The old CNN models that have increased depth, width or image size without a clear method. The use of neural architecture and a balanced scaling method introduces the EfficientNet family models. EfficientNet models achieve high accuracy by changing some parameters and FLOPs, which makes them suitable for transfer learning and more resources(Agustiani et al. 2025)(Mulim and Revikasha 2021). The efficientNet-Bo model uses limited resources and a small baseline and builds a large version of efficientNet on it. Given a baseline network with depth d_0 , width w_0 , and resolution r_0 , EfficientNet scales the model as:

$$d = d_0 \alpha^\phi, w = w_0 \beta^\phi, r = r_0 \gamma^\phi$$

Where $\phi \geq 0$ is the compound scaling coefficient and $\alpha, \beta, \gamma > 1$ are constants. Determined by the grid search. The scaling factors satisfy the computational constraint:

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$

Ensuring that each increment of \emptyset approximately doubles the computational cost. This balanced scaling approach helps EfficientNet models to achieve high accuracy-efficiency.

3.2.1 EfficientNetV2 faster training and fused operators

The author built an EfficientNetV2 on the EfficientNet series to build faster training and more efficient parameter use in the model. The author applies the progressive learning strategies in training aware NAS, adds fused MBConv blocks, such as resizing the image and adaptive regularization. These changes help to learn a model while training and improve accuracy (Ayman et al. 2024). EfficientNetV2-S, the mini version, trains very fast and achieves high accuracy with a small number of parameters. These models are good for experiments and training because the cost and results are essential.

Standard MBConv Block: This consists of three sequential operations:

$$F_{MBConv}(x) = Conv1 \times 1 \text{expandDWConv}k \times kConv1 \times 1 \text{reduce}x$$

Where x is the input feature map, DWConv denotes depthwise convolution, k is the kernel size. **Fused MBConv Block** EfficientNetV2 replaces the expansion and depthwise steps a single fused convolution.

$$F_{Fused}(x) = Convk \times kConv1 \times 1x$$

In EfficientNet and ResNet, a residual connection is applied.

$$y = x + F(x)$$

Ensures that stable gradient propagation and faster convergence.

3.3 ShuffleNetV2 and practical efficiency principles

When many models are developed, focus on real-world latency, optimizing FLOPs that are affected by operator overhead, memory access and specific platform. ShuffleNetV2 works on the actual speed of the system. They use a practical approach by using channel splitting and shuffling, and also follow the design guidelines (Wijaya et al. 2022). This model is a baseline, so it is easy to run on mobile and small machines. That uses the channel Split Operation and the given input feature map:

$$Y = Shuffle([X_1, Conv1 \times 1(DWConv3 \times 3(Conv1 \times 1X2))])$$

In ShuffleNetV2, the feature is grouped and split into two groups. One (X_1) is passed directly, while the other (X_2) is processed through pointwise and depthwise convolutions. The resulting concatenation is then shuffled to allow the exchange of information across channels. This design minimizes the memory and cost, and also improves computational efficiency in real hardware.

3.4 Comparative studies and benchmarks of architectures

In many studies compared, efficient CNNs like ShuffleNet, MobileNet, the EfficientNet family and EfficientNetV2 across different datasets and deployment settings. These studies rank the models its performance like heavily on the metric, and also care about such as accuracy, latency, training time, or energy use. They also use transfer learning for small models to boost significant performance (Hasan et al. 2024). No single model is best for every situation. EfficientNet and EfficientNetV2 achieve high accuracy and parameter efficiency (Ayman et al. 2024). Where memory limits are small or on low-latency machines, MobileNet or ShuffleNet is best. While there are some studies that achieved good results in waste image classification, the following problems are yet to be solved. Current methods focus either on classification performance and neglect computational efficiency, scalability, speed of inference, and applicability to real-world applications. Many studies are comparing different methods with different data sets and different training setups, which can make it hard to be sure how to evaluate the results to the best of one's ability. The restrictions emphasize the importance of developing a lightweight and computation-efficient hybrid system with high accuracy in classification, and also deploy in limited resource systems.

4. Methodology

This section presents the proposed HybridEffiShuffleNet (HES-Net) framework for automated waste image classification. The proposed architecture is designed to achieve an effective balance between classification accuracy

and computational efficiency for real-world smart waste management applications. HES-Net integrates EfficientNet-B0 and ShuffleNetV2 within a dual-branch architecture to exploit complementary feature representations while maintaining lightweight computational characteristics. The overall workflow of the proposed framework includes dataset preprocessing, dual-branch feature extraction, attention-based feature fusion, and final classification.

4.1 Overall Architecture of HES-Net

The proposed HybridEffiShuffleNet (HES-Net) architecture consists of two parallel feature extraction branches based on EfficientNet-B0 and ShuffleNetV2. The input waste image is resized to $224 \times 224 \times 3$ and simultaneously passed through both branches. EfficientNet-B0 is utilized to extract high-level semantic features using compound scaling, whereas ShuffleNetV2 captures lightweight structural and texture-based representations with low computational overhead.

The extracted feature maps from both branches are processed using Global Average Pooling (GAP) to reduce spatial dimensions and generate compact feature vectors. These feature vectors are concatenated and passed through an attention-based feature fusion module that adaptively assigns importance weights to discriminative features. Finally, the fused representation is forwarded to fully connected layers, followed by a Softmax classifier for multi-class waste categorization. The overall architecture of the proposed HES-Net framework is illustrated in Figure.

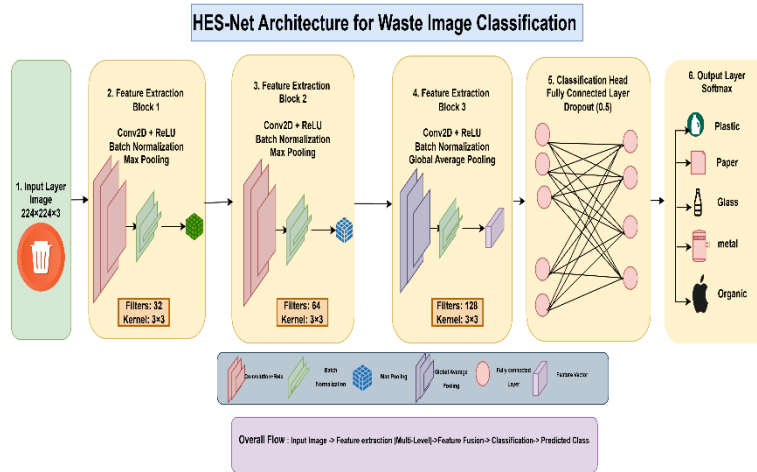


Fig.1: Architecture of the proposed HybridEffiShuffleNet (HES-Net) for automated waste classification.

4.2 EfficientNrt-BO Branch

EfficientNet-B0 serves as the primary semantic feature extraction branch in the proposed framework. EfficientNet utilizes a compound scaling strategy that uniformly scales network depth, width, and input image resolution to achieve improved performance with fewer parameters and reduced computational complexity. This balanced scaling mechanism enables EfficientNet-B0 to provide strong feature extraction capability while maintaining parameter efficiency. The compound scaling formulation used in EfficientNet can be expressed as:

$$y = \text{Softmax} \left(F_{cls} \left(\prod_{l=1}^L [X_l + M_{\alpha, \beta, \gamma} \phi(X_l)] \right) \right)$$

Explanation of the mathematical notation and how they work:

x_l : input feature mapping at layer l

L: Number of MBConv blocks

$M(\bullet)$: MBConv define as

$$M(X) = PWConv \left(SE \left(DWConv(PWConv(X)) \right) \right)$$

+: \$ residual connection.

$\alpha^\theta, \beta^\theta, \gamma^\theta$: That shows the Compound Scaling Factors for depth, width, and resolution

F_{cls} : Global average Pooling and a fully connected layer

Softmax: probability of final class estimation EfficientNet-B0 effectively captures high-level semantic information from waste images, enabling accurate classification of visually complex waste categories.

4.3 ShuffleNetV2 Branch

ShuffleNetV2 is incorporated into the proposed architecture to provide lightweight and computationally efficient feature extraction. ShuffleNetV2 employs channel splitting and channel shuffling operations to reduce memory access cost and computational complexity while maintaining strong feature learning capability.

The ShuffleNetV2 operation can be represented as

$$y = Softmax \left(F_{cls} \left(\prod_{l=1}^L S(C(X_l)) \right) \right)$$

Explanation of symbols

x_l : Input feature mapping at layer l

L: Total number of ShuffleNetV2 blocks

$C(\bullet)$: transformation function and split channel

$S(\bullet)$: channel shuffle operation

+: Identify (skip) connection.

F_{cls} : Global average Pooling and a fully connected layer

Softmax: class probability estimation ShuffleNetV2 significantly reduces computational overhead and inference latency, making it suitable for deployment in resource-constrained environments such as IoT devices and embedded systems.

4.4 Attention-Based Feature Fusion

To effectively combine complementary feature representations from EfficientNet-B0 and ShuffleNetV2, an attention-based feature fusion mechanism is introduced in the proposed HES-Net architecture. The feature vectors extracted from both branches are first concatenated and then passed through an attention module that adaptively assigns channel-wise importance weights to discriminative features.

The fused feature representation can be expressed as:

$$\hat{y} = Softmax(W_c(A \odot [GAP(f_s(X)); GAP(f_e(X))]))$$

where:

- $f_s(X)$ represents ShuffleNetV2 feature extraction,
- $f_e(X)$ represents EfficientNet-B0 feature extraction,
- GAP denotes Global Average Pooling,
- A represents attention weights,
- \odot denotes element-wise feature fusion,
- W_c represents the classification layer.

The attention mechanism enables the model to emphasize informative features while suppressing irrelevant feature responses, thereby improving classification robustness for visually similar waste categories.

4.5 Classification Layer

After attention-based feature fusion, the fused feature vector is forwarded to fully connected layers for final classification. A Softmax activation function is employed in the output layer to generate probability distributions across the five waste categories, namely glass, metal, organic, paper, and plastic.

The Softmax function is defined as:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}}$$

where:

- z_i denotes the output score of the class i ,
- C represents the total number of classes.

The final predicted class corresponds to the category with the highest probability score.

5. Dataset Preparation and Preprocessing

The waste image dataset used in this study consists of five categories: glass, metal, organic, paper, and plastic waste. The dataset was collected from publicly available online sources, and additional images were manually added to improve class diversity and dataset balance. The dataset was divided into training, validation, and testing subsets using a 70% / 15% / 15% split ratio to ensure fair model evaluation. All images were resized to 224×224 pixels to maintain compatibility with the input requirements of the CNN architectures. To improve model generalization and reduce overfitting, several data augmentation techniques were applied to the training set, including:

- Random horizontal flipping.
- Random rotation.
- Random cropping.
- Brightness and contrast adjustment.
- Image normalization.

The validation and testing datasets were not augmented to preserve unbiased evaluation conditions.

(a)



(b)



(c)

(d)



(e)



Fig.2: Sample images from the waste dataset illustrating five categories: (a) glass, (b) metal, (c) organic, (d) plastic, and (e) paper waste.

5.1 Training Configuration

All experiments were conducted under identical training settings to ensure fair comparison among the evaluated models. The models were trained using the Adam optimizer with categorical cross-entropy loss.

The training configuration used in this study is summarized in Table:1.

Table:1: Training Configuration of the Proposed HES-Net Framework

| Parameter | Value |
|-------------------|--------------------------------|
| Input Image Size | 224 × 224 |
| Backbone Networks | ShuffleNetV2 + EfficientNet-B0 |
| Feature Fusion | Concatenation |
| Optimizer | Adam |
| Learning Rate | 0.0003 |
| Weight Decay | 1×10^{-4} |
| Batch Size | 16 |
| Number of Epochs | 20 |

| | |
|-------------------------|---|
| Loss Function | Cross-Entropy Loss |
| Activation Function | ReLU / Softmax |
| Learning Rate Scheduler | StepLR (step size = 10, gamma = 0.1) |
| Dropout Rate | 0.3 |
| Input Normalization | ImageNet Mean and Standard Deviation |
| Data Augmentation | Horizontal Flip, Rotation (15°), Color Jitter |
| Hardware Platform | GPU-based System |

5.2 Computational Efficiency Analysis

In addition to classification accuracy, computational efficiency is an important consideration for real-world deployment. Therefore, the proposed HES-Net framework is evaluated using both performance and computational metrics, including:

- Classification Accuracy.
- Precision.
- Recall.
- F1-Score.
- Number of Parameters.
- Inference Time.
- Model Size.

The integration of EfficientNet-B0 and ShuffleNetV2 enables the proposed framework to achieve improved classification performance while maintaining computational efficiency suitable for practical smart waste management systems.

5.3 Workflow of the Proposed Framework

The overall workflow of the proposed HES-Net framework is summarized as follows:

1. Collection and preprocessing of waste image dataset.
2. Data augmentation and normalization.
3. Dual-branch feature extraction using EfficientNet-B0 and ShuffleNetV2.
4. Attention-based feature fusion.
5. Multi-class classification using fully connected and Softmax layers.
6. Performance evaluation using classification and computational metrics.

The proposed methodology aims to provide an efficient, scalable, and lightweight deep learning solution for automated waste image classification in real-world intelligent waste management applications.

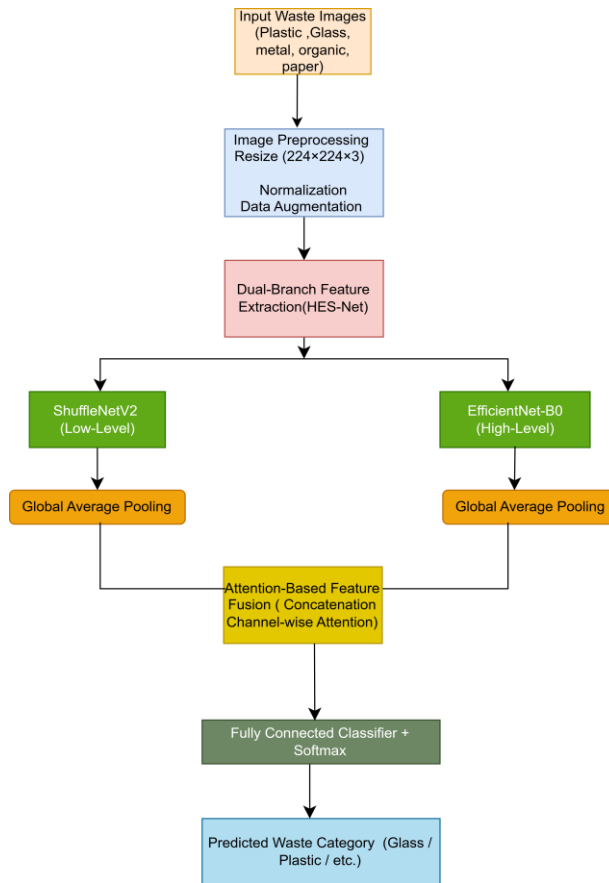


Fig.3: Workflow of the proposed HES-Net architecture for waste image classification using hybrid feature extraction and attention-based fusion.

6. Result and Discussion

This section presents the experimental evaluation and comparative analysis of the proposed HybridEffiShuffleNet (HES-Net) framework for multi-class waste image classification. The proposed model is evaluated against several state-of-the-art deep learning architectures, including ResNet-18, EfficientNet-B0, ShuffleNetV2, and ShuffleNetV2-S. All models were trained and evaluated under identical experimental conditions to ensure fair performance comparison. The evaluation was performed using multiple classification metrics, including Accuracy, Precision, Recall, and F1-Score.

6.1 Performance Evaluation Metrics

To comprehensively evaluate the performance of the proposed framework, several standard classification metrics were employed. Accuracy measures the overall classification correctness, while Precision, Recall, and F1-Score provide a more detailed assessment of classification performance across different waste categories.

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

where:

- TP denotes True Positives,
- TN denotes True Negatives,
- FP denotes False Positives,

- *FN* denotes False Negatives.

Precision, Recall, and F1-score are defined as follows

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

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

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

These metrics provide a balanced evaluation of the proposed framework across all waste categories.

6.2 Comparative Performance Analysis

The proposed HES-Net framework was experimentally compared with ResNet-18, EfficientNet-B0, ShuffleNetV2, and ShuffleNetV2-S using identical training configurations. Table 2 presents the comparative performance analysis of all evaluated models.

Table.2: Comparative Performance Analysis of Deep Learning Models

| Model | Accuracy | Precision | Recall | F1-Score |
|--------------------------------------|----------|-----------|--------|----------|
| ResNet-18 | 91.48% | 91.50% | 91.48% | 91.48% |
| EfficientNet-B0 | 95.20% | 95.20% | 95.20% | 95.20% |
| ShuffleNetV2 | 94.05% | 94.09% | 94.05% | 94.06% |
| ShuffleNetV2-S | 76.05% | 76.53% | 76.05% | 76.04% |
| HES-Net Without Attention (Proposed) | 96.53% | 96.54% | 96.53% | 96.53% |
| HES-Net With Attention (Proposed) | 96.63% | 96.64% | 96.63% | 96.63% |

Table 2: Provide a comparative study of various Deep learning models that were assessed for waste image classification. The accuracy of the baseline models like ResNet-18, 91.48%, EfficientNet-B0, 95.20% and shuffleNetV2 94.05% respectively, shuffleNetV2-S demonstrated comparatively lower performance with an accuracy of 76.05%, indicating limited feature extraction capability for the considered dataset. The proposed HES-Net architecture outperformed all baseline models. The HES-Net models achieve an accuracy of 96.53% without attention, demonstrating the effectiveness of hybrid feature fusion between ShuffleNet and Efficient-B0. Additionally, the integration of the attention mechanism improved the overall classification performance, with the highest accuracy of 96.63% along with superior precision, recall, and F1-Score values of 96.64%, 96.63% and 96.63%, respectively. The results show that the proposed attention-based fusion techniques enhance discriminative feature representation by highlighting more informative features during the classification process. Although the numerical improvement introduced by the attention mechanism appears relatively small, it is significant considering the already high baseline performance above 96%, where additional improvements become more difficult. Overall, the results of the experiments confirm that the suggested HES-Net architecture is resilient and effective for waste image classification tasks.

6.3 Training and Validation Analysis

The training and validation loss and accuracy curves of the proposed hybrid ShuffleNetV2–EfficientNet-B0 model are shown in Fig. 4. The training and validation losses decrease gradually over the training process, demonstrating successful feature learning and steady convergence. The training loss drops from 0.64 to 0.03, and the validation loss drops from 0.32 to 0.13, suggesting that the model is effectively reducing the number of classification errors on the training and validation sets. fig.4 (b) shows the accuracy curves for that. The training accuracy improves steadily from around 77% to 99%, while the accuracy of the validation improves from 89% to about 96%. Both curves

show an increasing behavior, which is good evidence for the successful learning of discriminative features in the waste image classification using the proposed architecture. There is a slight difference between the training and validation accuracy for the final training epochs. In particular, the training accuracy is still increasing after epoch 10, whereas the validation accuracy is converging around 96%. This behaviour is a sign of slight overfitting; nevertheless, the percentage difference is relatively small (around 3%), indicating that the model has decent generalisation capability. Moreover, the validation loss does not experience substantial changes, indicating that the model is not prone to overfitting. In general, the loss and accuracy curves validate that the proposed hybrid model could converge in an effective manner, learn the data well, and provide good generalization performance, which is suitable for accurate waste image classification.

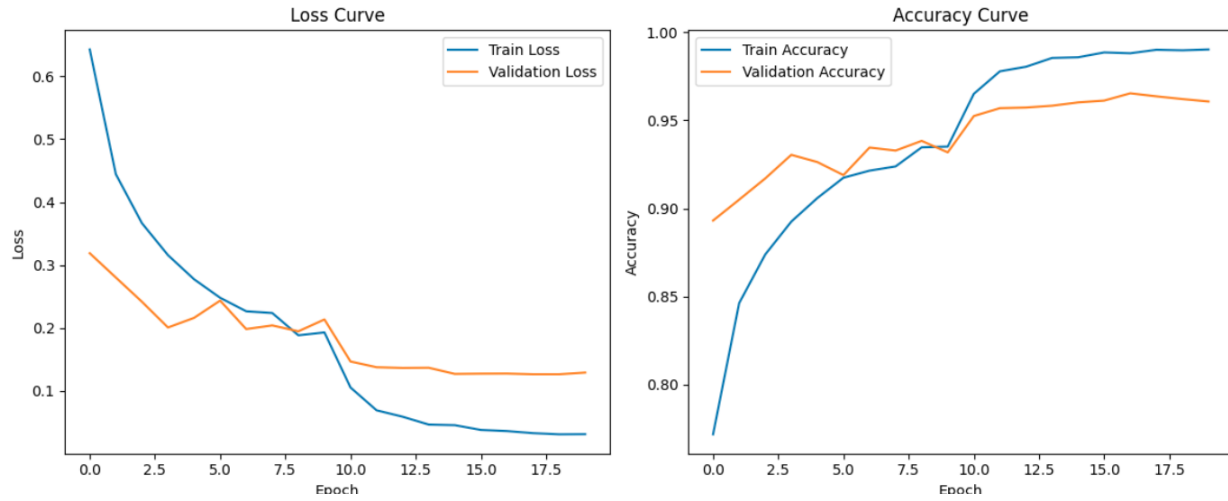


Fig. 4: Training and validation performance of the proposed hybrid model.

6.4 Confusion Matrix Analysis

The confusion matrices provide detailed insights into the classification behavior of the evaluated models. The proposed HES-Net framework exhibited strong diagonal dominance in the confusion matrix, indicating accurate classification across all waste categories with minimal inter-class confusion. EfficientNet-B0 also demonstrated strong classification capability; however, minor confusion was observed between paper and plastic categories. ShuffleNetV2 achieved reasonable performance but exhibited additional misclassifications in visually overlapping categories. ResNet-18 showed comparatively higher confusion between glass and metal categories because of limited fine-grained feature discrimination. ShuffleNetV2-S demonstrated the weakest confusion matrix performance due to its highly lightweight architecture and reduced representational capacity. These findings confirm that the proposed HES-Net framework effectively improves feature discrimination and classification consistency across multiple waste classes.

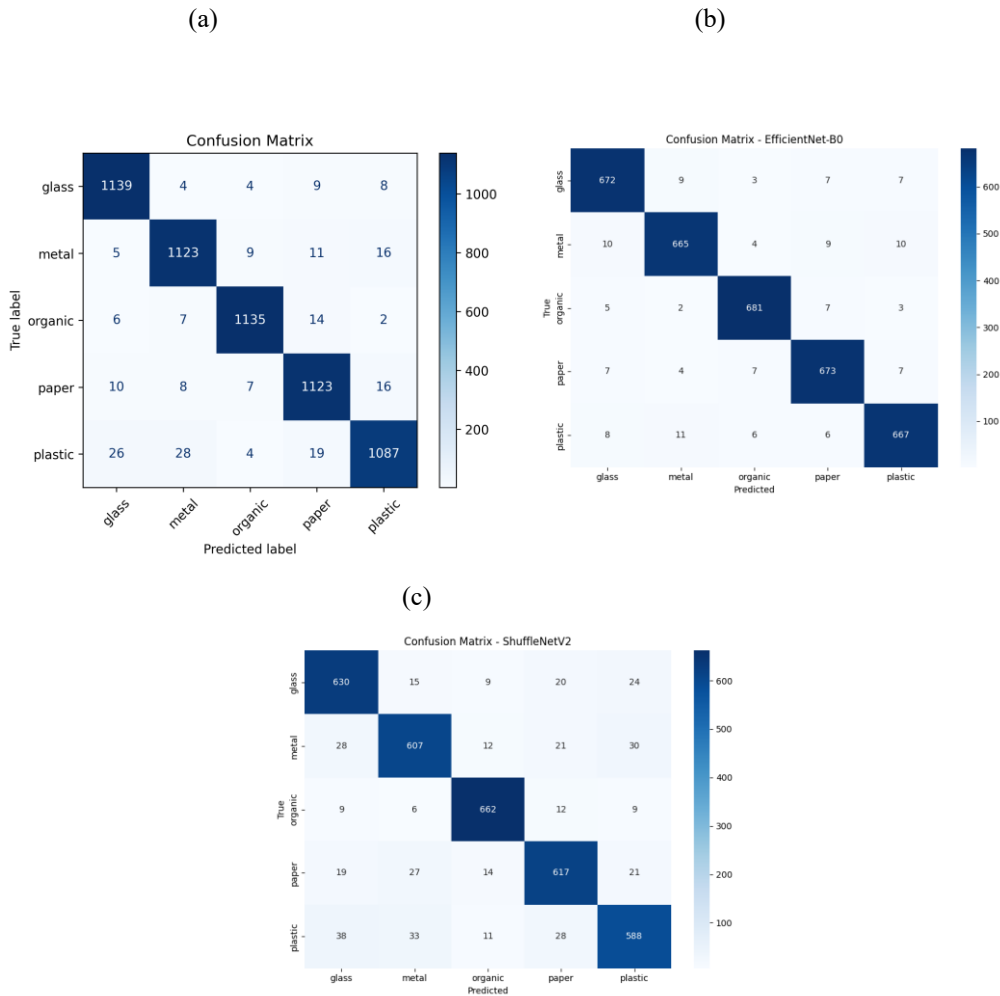


Fig.5: Confusion matrices of (a) HES-Net, (b) EfficientNet-B0, and (c) ShuffleNetV2.

6.5 Computational Efficiency Analysis

In addition to classification performance, computational efficiency is an important factor for real-world deployment in embedded systems, mobile devices, and IoT-based smart waste management platforms. Therefore, the proposed framework was evaluated using computational metrics such as parameter efficiency, model size, and inference capability. ShuffleNetV2 and ShuffleNetV2-S demonstrated strong computational efficiency because of their lightweight architecture design. EfficientNet-B0 achieved improved accuracy but required comparatively higher computational resources. The proposed HES-Net framework successfully balanced classification accuracy and computational efficiency by integrating EfficientNet-B0 and ShuffleNetV2 within a unified hybrid architecture. The experimental findings indicate that HES-Net provides a practical trade-off between performance and efficiency, making it suitable for real-world intelligent waste classification applications where both accuracy and lightweight deployment are essential.

6.6 GradCam

In order to make the performance of the proposed HES-Net more interpretable and to study the decision-making characteristics of the proposed HES-Net framework, Gradient-weighted Class Activation Mapping (Grad-CAM) was used. Grad-CAM creates explanations as visual markers on the image that contribute most to the model prediction. Figure 5 shows the Grad-CAM visualizations for the various waste types, such as glass, metal, organic, paper, and plastic. The heat maps obtained through the proposed framework are shown to highlight relevant object regions and minimise irrelevant background information. The obtained visualization results show the improvement of localization of features and the interpretability of classification by using the attention-based feature fusion mechanism. Specifically, the proposed framework is able to incorporate valuable semantic and structural properties of waste classes visually similar to each other, like paper and plastic. The results show the potential of HES-Net to make accurate and interpretable predictions for smart waste management.

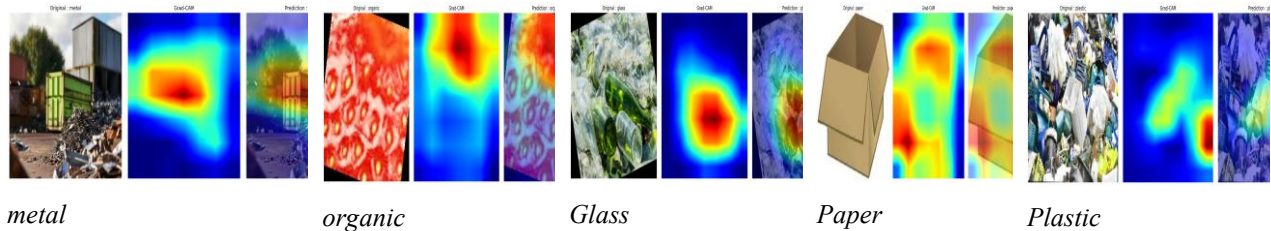


Fig.6: Grad-CAM visualizations of the proposed HES-Net framework for different waste categories. The model effectively focuses on discriminative object regions for accurate classification.

6.7 Discussion

The experimental results confirm that hybrid feature extraction strategies can significantly improve waste image classification performance compared with individual CNN architectures. The proposed HES-Net framework is able to effectively combine the semantic feature extraction capability of EfficientNet-B0 with the lightweight structural representation learning capability of ShuffleNetV2. The attention-based feature fusion mechanism also improves the generalization performance of the classification in an adaptive way by emphasizing the discriminative feature representations. The proposed model to attain better classification accuracy while preserving the computational efficiency also they need to practical development. The framework proposed was successful, but had some drawbacks. The present study was assessed using a five-class waste image database with minimal environmental variations. The data sets could be expanded with larger and more diverse data sets, real-time deployment analysis, transformer-based architectures, and the use of explainable AI (XAI) methods like Grad-CAM and attention visualization for improved interpretability in future work. Overall, the proposed HES-Net concept can be successfully employed in the development of the smart waste management system and automated smart waste picking system.

7. Conclusion and Future Work

In this paper, a lightweight hybrid deep learning framework is proposed for automated waste image classification dubbed as HybridEffiShuffleNet (HES-Net). To achieve the efficient complementary semantic and structural features from the waste images, a model named EfficientNet-B0+ShuffleNetV2 with an attention-based feature fusion mechanism was proposed. Experimental tests showed that HES-Net can outperform ResNet-18, EfficientNet-B0, ShuffleNetV2 and ShuffleNetV2-S with the same training method and comparable accuracy of 96.34%, in a five-class waste image dataset. The proposed framework was also found to have a good classification consistency over all waste categories and was also found to be computationally efficient, suitable for smart waste management applications. These results show that HES-Net is a good compromise between accurate classification and deployment simplicity for in-the-wild intelligent recycling systems. Future research will be conducted on larger real-world datasets, Transformer-based architectures, Explainable AI techniques, and real-time deployment optimisation of edge-based smart waste management systems.

References

1. Agustiani, Sarifah, Agus Junaidi, Riska Aryanti, and Anton Abdul Basah Kamil. 2025. "Waste Classification Using EfficientNetB3-Based Deep Learning for Supporting Sustainable Waste Management." *Bulletin of Informatics and Data Science* 4 (1): 62–70.
2. Ahmad, Iman Ameer, Ahmed Mudheher Hasan, and Amjad Jaleel Humaidi. 2025. "Development of a Memory-Efficient and Computationally Cost-Effective CNN for Smart Waste Classification." *Journal of Engineering Research*.
3. Ahmed, Mohammed Imran Basheer, Raghad B Alotaibi, Rahaf A Al-Qahtani, et al. 2023. "Deep Learning Approach to Recyclable Products Classification: Towards Sustainable Waste Management." *Sustainability* 15 (14): 11138.
4. Alalibo, TAMUNO-OMIE J, and NKOLIKA O Nwazor. 2023. "Comparative Analysis of Convolutional Neural Network Models for Solid Waste Categorization." *Iconic Research and Engineering Journals* 6 (12): 1396–405.
5. Almtireen, Natheer, Abdelrazzaq A Abuhejleh, Mutaz Ryalat, Hisham Elmoaqet, and Ghaith Al-refai. 2025. "Deep Transfer Learning for Sustainable Waste Management: Real-Time Waste Segregation Apparatus Using a Two-Phase CNN Framework: N. Almtireen Et Al." *The Journal of Supercomputing* 81 (15): 1411.
6. Ayman, Umme, Mohammad Asifur Rahim, Ireen Ara Haque, Md Mizanur Rahman, Drisrty Saha, and Md Azamain Mahtab Rahat. 2024. "EfficientNet-Based Deep Learning Model for Advanced Waste Classification." *2024 27th International Conference on Computer and Information Technology (ICCIT)*, 2564–69.
7. Dawood, Afrah Salman. 2023. "A Comparative Study of DCNN Models and Transfer Learning Effect for Sustainability Assessment: The Case of Garbage Classification." *Technium* 12.
8. Eurostat. 2026. *Waste Statistics*. Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Waste_statistics.
9. Gurcan, Fatih, and Ahmet Soylu. 2025. "Automated Waste Classification for Smart Recycling: A Multi-Class CNN Approach with Transfer Learning and Pre-Trained Models." *Environmental Technology & Innovation*, 104673.
10. Hasan, SM Mahedy, Md Farukuzzaman Faruk, Azmain Yakin Srizon, Md Ali Hossain, and Md Al Mamun. 2024. "Cost-Sensitive Lightweight Waste Classification Using Modified EfficientNetB0 and Parallel CNNs for Resource-Constrained Environments." *2024 27th International Conference on Computer and Information Technology (ICCIT)*, 1212–17.
11. Ji, Tianchen, Huaiying Fang, Rencheng Zhang, Jianhong Yang, Lulu Fan, and Jiantao Li. 2023. "Automatic Sorting of Low-Value Recyclable Waste: A Comparative Experimental Study." *Clean Technologies and Environmental Policy* 25 (3): 949–61.
12. Lilhore, Umesh Kumar, Sarita Simaiya, Surjeet Dalal, and Robertas Damaševičius. 2024. "A Smart Waste Classification Model Using Hybrid CNN-LSTM with Transfer Learning for Sustainable Environment." *Multimedia Tools and Applications* 83 (10): 29505–29.
13. Lin, Kunsen, Youcai Zhao, Lina Wang, Wenjie Shi, Feifei Cui, and Tao Zhou. 2023. "MSWNet: A Visual Deep Machine Learning Method Adopting Transfer Learning Based Upon ResNet 50 for Municipal Solid Waste Sorting." *Frontiers of Environmental Science & Engineering* 17 (6): 77.
14. Mulim, W., and M. F. Revikasha. 2021. "Waste Classification Using EfficientNet-B0." *Proceedings of the 1st International Conference on Artificial Intelligence and Computer Science*. <https://ieeexplore.ieee.org/document/9609756>.
15. Van de Schepop, A. 2025. *A Comparative Analysis of Waste Image Classification: Xception Vs EfficientNet-B0*. Online. <http://arno.uvt.nl/show.cgi?fid=175734>.
16. Wijaya, Eka Setya, Andy Mizwar, Achmad Mujaddid Islami, Yuslena Sari, Erika Maulidiya, and Irham Maulani Abdul Gani. 2022. "Garbage Classification Using CNN Architecture ShuffleNet V2." *2022 Seventh International Conference on Informatics and Computing (ICIC)*, 1–4.
17. World Bank. 2026. *What a Waste*. World Bank Data Topics. <https://datatopics.worldbank.org/what-a-waste/>.