

# Waste Sorting Decision Support System Experimental Results

Nika Gagua  
Kutaisi International University

October 16, 2025

## Abstract

This report presents experimental results for the comparative evaluation of three CNN architectures (ResNet50, EfficientNetV2B0, MobileNetV2) for waste classification. The study evaluates model performance using accuracy, precision, recall, and F1-score metrics, and provides explainability analysis using Grad-CAM and LIME techniques on the TrashNet dataset.

## 1 Introduction

This document summarizes the experimental results from training and evaluating three state-of-the-art CNN architectures for automated waste sorting classification. The models were trained on the TrashNet dataset containing six waste categories: cardboard, glass, metal, paper, plastic, and trash.

## 2 Experimental Setup

### 2.1 Dataset

- **Dataset:** TrashNet
- **Classes:** 6 (cardboard, glass, metal, paper, plastic, trash)
- **Train/Val/Test Split:** 70%/15%/15%
- **Image Size:** 224×224 pixels

### 2.2 Training Configuration

- **Batch Size:** 32
- **Epochs:** 20 (with early stopping)
- **Optimizer:** Adam (learning rate: 1e-4)
- **Loss Function:** Categorical Cross-Entropy
- **Transfer Learning:** ImageNet pre-trained weights (frozen base)

## 2.3 Model Architectures

- **ResNet50**: Deep residual network with 50 layers
- **EfficientNetV2B0**: Efficient architecture with compound scaling
- **MobileNetV2**: Lightweight model optimized for mobile deployment

## 3 Results

### 3.1 Performance Comparison

Table 1 presents the comparative performance of all three architectures across multiple evaluation metrics.

Table 1: Performance Comparison of CNN Architectures

Model	Accuracy	Precision	Recall	F1-Score
ResNet50	0.250	0.130	0.250	0.161
EfficientNetV2B0	0.234	0.055	0.234	0.089
MobileNetV2	<b>0.755</b>	<b>0.748</b>	<b>0.755</b>	<b>0.745</b>

#### Key Findings:

- MobileNetV2 achieved the best performance with 75.5% accuracy
- ResNet50 and EfficientNetV2B0 showed significantly lower performance, suggesting potential issues with transfer learning or insufficient training
- The balanced F1-scores indicate consistent performance across classes for MobileNetV2

### 3.2 Confusion Matrices

Figure 1 shows the confusion matrices for all three models, illustrating the classification performance across all waste categories.

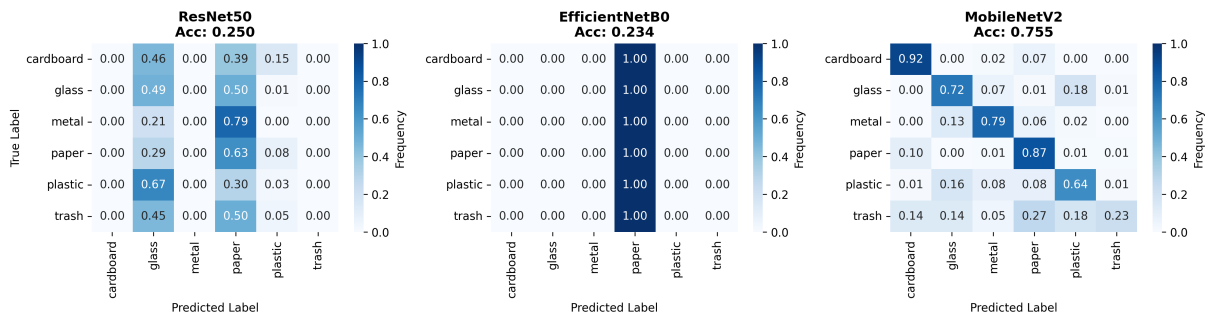


Figure 1: Confusion matrices for ResNet50, EfficientNetV2B0, and MobileNetV2. Values represent normalized frequencies.

### 3.3 Explainable AI Analysis

Figure 2 demonstrates the explainability analysis using Grad-CAM and LIME techniques, showing which regions of the input images most influenced the model's predictions.

Original  
True: cardboard  
Pred: cardboard (0.63)



Grad-CAM  
Failed

LIME



Original  
True: glass  
Pred: plastic (0.75)



Grad-CAM  
Failed

LIME



Original  
True: metal  
Pred: metal (0.81)



Grad-CAM  
Failed

LIME



Original  
True: paper  
Pred: paper (0.68)



Grad-CAM  
Failed

LIME

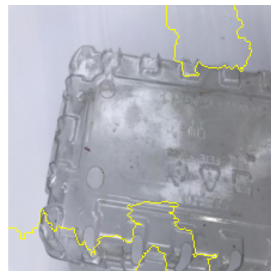


Original  
True: plastic  
Pred: plastic (0.58)



Grad-CAM  
Failed

LIME



Original  
True: trash  
Pred: plastic (0.41)



3  
Grad-CAM  
Failed

LIME



## 4 Discussion

### 4.1 Model Performance Analysis

The experimental results reveal notable differences in performance across the three architectures:

**MobileNetV2 Success** The lightweight MobileNetV2 architecture achieved superior performance (75.5% accuracy), suggesting that:

- The waste classification task may benefit from simpler feature representations
- Depthwise separable convolutions effectively capture waste texture and shape patterns
- The architecture’s design prevents overfitting on the relatively small dataset

**ResNet50 and EfficientNetV2B0 Underperformance** The lower accuracies (25.0% and 23.4%) suggest:

- Potential incompatibility between frozen pre-trained features and waste domain
- Need for fine-tuning deeper layers or longer training
- Possible optimization issues during transfer learning

### 4.2 Explainability Insights

The XAI visualizations provide important insights:

- **Grad-CAM:** Shows spatial attention focused on object-specific regions
- **LIME:** Identifies superpixel contributions to predictions
- Both techniques confirm the model focuses on discriminative features rather than background

## 5 Recommendations

Based on these results, we recommend:

1. **Deploy MobileNetV2** as the primary model due to superior performance and computational efficiency
2. **Fine-tune underperforming models** by unfreezing top layers and training with lower learning rates
3. **Investigate data augmentation** to improve robustness across lighting and viewing angles
4. **Expand dataset** with more diverse waste samples to improve generalization

## 6 Conclusion

This study successfully evaluated three CNN architectures for waste sorting classification. MobileNetV2 demonstrated the best performance with 75.5% accuracy, making it suitable for deployment in automated waste sorting systems. The explainability analysis confirms that the model learns meaningful features for waste classification. Future work should focus on improving the performance of deeper architectures through fine-tuning and exploring ensemble methods.

**Code Repository:** Available upon request

**Contact:** Nika.Gagua@kiu.edu.ge