EXPLAINABLE AI · DECISION SUPPORT · SUSTAINABLE WASTE MANAGEMENT

An Explainable AI-Based Decision Support System for Waste Sorting Systems

Ia Mosashvili[1] · Nika Gagua[1] · Nikolaos Pagonis[2] · Konstantinos Kokkinos[3] · Konstantinos Moustakas[2] · Vayos Karayannis[4]*
[1] Computer Science School, Kutaisi International University, Georgia · [2] Digital Systems Dept., University of Thessaly, Greece
[3] School of Chemical Engineering, National Technical University of Athens, Greece · [4] Chemical Engineering Dept., University of Western Macedonia, Greece

Deep Learning · XAI · Waste Classification · Decision Support · Transfer Learning · Circular Economy

MOTIVATION

Municipal solid waste (MSW) is projected to reach **3.4 billion tons/year by 2050**. Convolutional neural networks (CNNs) excel at waste image classification but remain opaque "black-box" systems — a critical barrier for regulatory compliance and operator trust in real-world sorting facilities. This work bridges that gap with an Explainable Decision Support System (X-DSS).

OBJECTIVES

① Design CNN classifiers on TrashNet benchmark
② Integrate multiple XAI techniques for transparent model reasoning
③ Align with sustainability policy and industrial accountability standards

DATASET — TRASHNET

Class	N
Paper	594
Glass	501
Plastic	482
Metal	410
Cardboard	403
Trash	137
Total	**2,527**

Images: 224×224 px · Normalized [0,1] · Augmentation on training set only (flip, rotation ±15°, zoom ±15°, contrast, brightness)

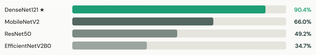
BEST MODEL — DENSENET121

90.41% Test Accuracy
0.896 Weighted F1
0.904 Precision
0.904 Recall

METHODOLOGY OVERVIEW

Four CNN architectures pre-trained on ImageNet were fine-tuned using two-phase transfer learning: (1) classification head training at $lr = 10^{-3}$ for 15 epochs; (2) full network fine-tuning at $lr = 10^{-4}$ for 35 further epochs. Class imbalance was addressed via balanced sample weighting. Four post-hoc XAI methods were applied to the best model to produce complementary, human-interpretable explanations.

CNN MODEL COMPARISON — TEST ACCURACY



DenseNet121's dense connectivity and feature reuse yield superior performance on small, moderately imbalanced datasets. EfficientNetV2B0 requires larger-scale data for its compound scaling to generalise.

XAI TECHNIQUES APPLIED

Grad-CAM
Gradient-weighted spatial heatmaps from the last convolutional layer highlight material-relevant regions (e.g., bottle edges, cardboard texture).

LIME
Local superpixel perturbations identify which image segments drive each classification, providing model-agnostic local explanations.

Integrated Gradients
Path-based pixel-level attribution from baseline to input reveals fine-grained cues: metallic specular edges, transparency gradients in plastics.

Occlusion Sensitivity
Systematic patch occlusion quantifies region importance — confirming model focus on material surfaces rather than background noise.

KEY INTERPRETABILITY FINDINGS

All four methods converged on consistent saliency regions: corrugated texture patterns for *cardboard*, specular reflections and crisp outlines for *metal*, printed text and flat edges for *paper*, container shape and transparency for *plastic*, and reflective curvature for *glass*. The *trash* class distributed attention across multiple regions, reflecting its inherent visual heterogeneity. Agreement between gradient-based (Grad-CAM, IG) and perturbation-based (LIME, OS) approaches validates the robustness of the X-DSS's reasoning.

LIMITATIONS

- Controlled dataset (2,527 images); limited real-world diversity
- Absent waste types: e-waste, organic, hazardous
- Glass class most challenging (56.6% per-class accuracy)
- 90.4% accuracy may require human oversight for high-stakes automation
- Unimodal (image-only); no density, NIR, or weight data

FUTURE DIRECTIONS

- Domain adaptation for real-world conveyors (varying illumination, occlusion)
- Expanded datasets including hazardous and organic streams
- Multimodal sensing: NIR, hyperspectral, weight/density
- Uncertainty quantification for human-in-the-loop triage
- Edge deployment with model compression for real-time inference

CONCLUSIONS

This study presents the first X-DSS that couples CNN-based waste classification with multi-faceted XAI in a unified, policy-aligned framework. DenseNet121 achieved 90.41% accuracy and a weighted F1 of 0.896 on the TrashNet benchmark — outperforming MobileNetV2, ResNet50, and EfficientNetV2B0. The combined application of Grad-CAM, LIME, Integrated Gradients, and Occlusion Sensitivity transforms opaque CNN decisions into auditable, human-understandable explanations, directly supporting regulatory compliance and operational trust in automated sorting facilities.

Data: TrashNet (Yang & Thung, 2016) · No funding declared · No competing interests

X-DSS · WASTE AI · XAI