International Journal of Computer Applications (0975 - 8887)
Volume 187 - No.116, June 2026

ReLeaf: A MobileNetV2-based Mobile Application for Real-Time Waste Classification with LLM-Assisted Recycling Guidance

Fatimah H. Alyami
College of Computer Science & IT
Imam Abdulrahman Bin Faisal University
Dammam, Saudi Arabia

Nadeen N. Abduljabbar
College of Computer Science & IT
Imam Abdulrahman Bin Faisal University
Dammam, Saudi Arabia

Ghadi T. Alzahrani
College of Computer Science & IT
Imam Abdulrahman Bin Faisal University
Dammam, Saudi Arabia

Dana B. Alakeel
College of Computer Science & IT
Imam Abdulrahman Bin Faisal University
Dammam, Saudi Arabia

Amal S. Almirsal
College of Computer Science & IT
Imam Abdulrahman Bin Faisal University
Dammam, Saudi Arabia

Atheer S. Algherairy
College of Computer Science & IT
Imam Abdulrahman Bin Faisal University
Dammam, Saudi Arabia

ABSTRACT

This paper introduces ReLeaf, an intelligent waste classification system that integrates deep learning and mobile application technologies to promote environmentally responsible waste disposal. The approach employs the MobileNetV2 architecture with transfer learning and fine-tuning to classify waste into six categories: cardboard, glass, metal, paper, plastic, and trash. Various data augmentation techniques were systematically evaluated to enhance model generalization and mitigate overfitting. The model was trained on a combination of public datasets and achieved a test accuracy of 93.96%. The trained model was deployed within a Flutter-based mobile application using TensorFlow Lite, enabling real-time waste recognition on mobile devices. Additionally, a cloud-based large language model (GPT-4.1-mini) was incorporated to provide recycling guidance and user assistance through natural language interaction. Experimental results indicate that the proposed system offers an efficient, practical solution for intelligent waste classification and recycling support in real-world settings.

General Terms

Deep Learning, Image Classification, Mobile Application

Keywords

Waste Classification, Deep Learning, MobileNetV2, Transfer Learning, Data Augmentation, TensorFlow Lite, Recycling, Smart Waste Management, Large Language Model

1. INTRODUCTION

Improper waste disposal and low recycling awareness continue to pose major environmental and sustainability challenges worldwide. With the rapid growth of urban waste and increasing environmental concerns, intelligent waste management systems have become an important research area. Recent advances in artificial intelligence, particularly deep learning and image classification, have enabled the development of automated waste recognition systems that can support recycling and waste sorting processes.

Despite the promising performance reported in previous studies, several challenges remain. Variations in lighting conditions, object shapes, backgrounds, and dataset diversity can negatively affect model generalization in real-world environments. In addition, many existing approaches primarily focus on improving classification accuracy through complex architectures. At the same time, limited attention has been given to systematically evaluating the effect of data augmentation techniques when using lightweight models suitable for mobile deployment.

Therefore, this study investigates the effectiveness of the MobileNetV2 architecture for waste classification via transfer learning and fine-tuning. Different augmentation configurations are evalu-

ated to improve model robustness, reduce overfitting, and maintain computational efficiency.
Furthermore, this work introduces *ReLeaf*, a mobile application designed to provide real-time waste classification and recycling assistance via an accessible, user-friendly interface.
The main objectives of this study are as follows:

—To review existing studies on AI-based waste classification and smart recycling systems.

—To evaluate the effectiveness of MobileNetV2, using transfer learning and fine-tuning techniques, for waste image classification.

—To investigate the impact of different data augmentation techniques on model performance and generalization.

—To design and develop a user-friendly mobile application (*ReLeaf*) that integrates a lightweight deep learning model for real-time waste classification.

The remainder of this paper is organized as follows: Section 2 reviews existing studies on AI-based waste classification and smart recycling systems. Section 3 presents the proposed classification model, including the dataset, preprocessing steps, image augmentation techniques, and experimental results. Section 4 describes the development of the *ReLeaf* system and highlights its main features. Finally, Section 5 concludes the paper.

2. LITERATURE REVIEW

2.1 Deep Learning Approaches

Recently, ”Deep Learning” has become the most widely used approach for automatic waste classification. Neural networks are trained directly on raw image data, eliminating the need for manual feature extraction and achieving higher accuracy than traditional methods. Various CNN architectures such as MobileNet, EfficientNet, ResNet, and DenseNet have demonstrated strong performance in identifying waste materials.
Lightweight models such as MobileNet are particularly suitable for low-powered and mobile devices. For example, [30] achieved 82.92% accuracy, while [8] achieved 86.73% using CPU-only systems. Preprocessing techniques such as resizing, normalization, and enhancement significantly improve performance, reaching up to 96% accuracy [1].
Data augmentation techniques such as rotation and flipping further enhance model performance. Studies such as [26] reported 95.4% accuracy, while other CNN-based approaches achieved accuracy above 90%, making them reliable for waste classification tasks.

2.2 Hybrid/ Advanced Deep Learning

Hybrid approaches combine CNNs with techniques such as attention mechanisms, transformers, and ensemble models to improve classification accuracy. These methods aim to leverage the strengths of multiple architectures.
For instance, [31] achieved 92% accuracy using attention mechanisms, while [5] combined ResNet50 and MobileNet to reach 97.9%. Transformer-based approaches such as [16] achieved up to 98.27%.
Ensemble methods also improve stability and reduce overfitting. For example, [17] combined multiple models and achieved around 97% accuracy.

2.3 Vision Language/ Zero-Shot / Multimodal Approaches

Multimodal approaches integrate vision and language models to enhance system capabilities beyond classification. These models can describe images, answer questions, and identify unseen categories.
For example, [32] combined CLIP with a language model to enable image understanding and text-based interaction. Zero-shot learning approaches, such as [22], allow models to classify unseen waste categories without additional training.

2.4 Mobile / Internet of Things (IoT) Application Implementation

Recent research has focused on deploying waste classification models on mobile and IoT devices. Lightweight architectures such as MobileNet and EfficientNet enable efficient processing with limited resources.
Mobile-based applications have achieved accuracy above 90%, such as [11] and [19]. Real-time detection using smartphones has also been demonstrated, achieving high performance in practical scenarios [4].
In addition, IoT-based systems integrate classification models into smart environments, enabling automated waste management solutions [9].

2.5 Summary and Research Gap

Despite the significant progress in deep learning-based waste classification, several limitations remain. Many existing studies focus primarily on achieving high accuracy using complex or computationally intensive models, which may not be suitable for real-time or resource-constrained applications. Additionally, while data augmentation is widely used, its impact is often not systematically evaluated across different configurations, particularly in combination with lightweight architectures such as MobileNetV2. Furthermore, many studies rely on relatively small or less diverse datasets, which limits the generalization of the models in real-world scenarios.
Therefore, this study aims to address these gaps by systematically evaluating the performance of MobileNetV2 using transfer learning and fine-tuning, while investigating the impact of different data augmentation techniques on model performance and generalization. In addition, the trained model is integrated into a mobile-based application to demonstrate its practical applicability in real-world waste classification scenarios.

A summary of the reviewed studies is presented in Table 1.

Table 1. : Summary of Reviewed Studies

Ref.	**Year**	**Dataset**	**Categories**	**Model type**	**Preprocessing**	**Task Type**	**Result**	
							Precision	**Accuracy**
Model type: MobileNet Family (V1 / V2)								
[19]	2023	TrashNet	6 Categories	MobileNetV2	Image resizing, normalization, and data augmentation (rotation, translation, scaling, brightness, shearing).	Multi-class classification	92%	92%
[31]	2023	TrashNet	6 Categories	MobileNetV2	Data augmentation, image resizing and normalization	Multi-class classification	-	92%
[8]	2025	TrashNet	6 Categories	MobileNetV2	Image resizing, normalization, and data augmentation (flip, rotation, caching/prefetching).	Multi-class classification	86%	86.73%
[13]	2024	Garbage classification dataset	6 Categories	MobileNetV2 combined with 4 other models	Fourier-based deblurring, curvelet-based empirical Wiener filtering, joint non-local means filtering, and reflection removal	Multi-class classification	96%	97%
[5]	2024	Kaggle and GitHub (Secondary)	12 Categories	MobileNetV combined with ResNet50	Data augmentation, image resizing, normalization and data splitting	Binary and multi-class classification	-	97.9%
[1]	2025	Custom	2 Categories	MobileNetV1	Image resizing, class balancing, and data preprocessing (CLAHE + Bilateral Filtering)	Binary classification	93%	96%
[30]	2023	Web And Public Sources	4 Categories	MobileNetV2	Image resizing, cleaning, and data splitting, normalization	Multi-class classification	96%	82.92%
[25]	2023	Custom	10 Categories	MobileNetV2	Image resizing, normalization, data splitting, data augmentation (ImageDataGenerator)	Multi-class classification	88%	88.64%
Model Type: DenseNet Family (DenseNet121 / DenseNet201)								
[7]	2025	TrashNet, CompostNet	7 Categories	DenseNet121	Image resizing, normalization, and data augmentation (rotation, flip, shear), and data down sampling.	Multi-class classification	-	97%
[17]	2025	Trashnet, Waste Classification V2, Waste Classification, Openrecycle	27 Categories	DenseNet201	Image resizing and normalization	classification	97%	97%
Model Type: EfficientNet Family (EfficientNet-Lite0 / EfficientDet-Lite3 / CE-EfficientNetV2)								
[11]	2022	Garbage Classification Dataset	11 Categories	EfficientNet-Lite0	Image resizing, normalization, and data augmentation (rotation, flip)	Multi-class classification	-	95.39%
[12]	2023	Garbage Classification Dataset	2 Categories	EfficientNet-Lite0	Image resizing and normalization	Binary classification	-	82%
	2023	Garbage Classification Dataset	11 Categories	EfficientDet-Lite3	Image resizing, bounding-box labeling, normalization, and data augmentation (rotation, enhancement)	Object detection	-	78%
[12]	2025	TrashNet	5 Categories	CE-EfficientNetV2	Image resizing, normalization, and data augmentation (flip, rotation, translation, noise injection).	Multi-class classification	-	96.5%
	2025	Huawei Cloud Waste Dataset	4 Categories	CE-EfficientNetV2	Image resizing, normalization, and data augmentation (flip, rotation, translation, noise injection).	Multi-class classification	-	95.4%
Model Type: ResNet-based Models (ResNet50 / ResNeXt-50)								
[29]	2024	TrashNet	6 Categories	ResNeXt-50	Image resizing, normalization, and data augmentation (crop, flip, color jitter, rotation)	Multi-class classification	99%	99.17%

Ref.	Year	Dataset	Categories	Model type	Preprocessing	Task Type	Result	
							Precision	Accuracy
[2]	2025	Garbage Dataset	12 Categories	ResNet50	Image resizing, noise reduction, normalization, and data augmentation (flip, rotation, contrast, SMOTE oversampling)	Multi-class classification	97.98%	98.16%
Model Type: Hybrid Models (ResNet + ViT / DenseViT / HR-ViT)								
[16]	2025	Custom	6 Categories	HR-ViT (Hybrid ResNet50 + ViT)	Image resizing, normalization, and data augmentation (rotation, shift, shear, zoom, flip)	Multi-class classification	98%	98.27%
[28]	2025	WaRP Dataset	28 Categories	2S_DenseViT (DenseNet-201 + MaxViT)	Image resizing, normalization, and data augmentation (histogram equalization, cropping/padding, color adjustment, horizontal, vertical flips, infill).	Classification + Detection	-	83.11%
Model Type: YOLO Family (YOLO-11n / YOLOv11x)								
[6]	2025	Trashnet	7 Categories	YOLOv11x	Image resizing, normalization, and data augmentation (shear, zoom, flip)	Multi-class classification	94.8%	94.8%
[20]	2024	Wadaba	5 Categories	YOLO-11n	Image resizing, normalization, oversampling, and data augmentation (random zoom, rotation, contrast adjustment, flip)	Multi-class classification	98%	95.53%
Model Type: Faster R-CNN Family (Object Detection CNNs)								
[4]	2023	Custom	4 Categories	Faster R-CNN	Image resizing, normalization, and data augmentation (enhancement, rotation)	Object detection	-	98.11%
[10]	2024	Garbage Classification Dataset	5 Categories	Faster R-CNN	Image resizing, normalization, and data augmentation (cleaning, rotation)	Object detection	-	78.65%
Model Type: Classical CNNs (Enhanced CNN / Deep CNN / VGG16 + Random Forest)								
[14]	2024	Trashnet	6 Categories	Deep CNN	Image resizing, normalization, and data augmentation Image resizing, normalization, and data augmentation (rotation, flip, zoom, brightness, contrast)	Multi-class classification	97%	97.3%
[18]	2023	Organic And Recyclable Waste Images	2 Categories	Enhanced CNN	Image resizing, normalization, cleaning, and data augmentation (crop, rotation, flip, brightness, noise)	Binary Classification	96%	94%
[9]	2025	Garbage Classification Dataset	6 Categories	VGG16 + Random Forest	Image resizing, normalization, label encoding, one-hot encoding, train-test split, and data augmentation (rotation, flip, zoom, brightness, contrast)	Image classification	-	85%
Model Type: Transformer-based Models (ViT / OpenCLIP / MiniGPT-4 / CLIP)								
[22]	2025	Trashnet	6 Categories	OpenCLIP ViT-L/14@336px (Zero-Shot)	Image resizing and normalization	Multi-class classification	-	76.30%
[15]	2021	Trashnet	6 Categories	ViT	Image resizing and normalization	Multi-class classification	-	96.98%
[21]	2025	Kumsetty et al. (2022)	6 Categories	OpenCLIP ViT-L/14-2B (BigG)	Image resizing, center-cropping, RGB conversion, tensor transformation, and normalization	Multi-class classification	-	97.18%
[32]	2023	LAION-400M, Conceptual Captions, SBU	-	MiniGPT-4 (ViT-G/14 + Vicuna LLM with Q-Former)	Image resizing and normalization, visual feature extraction using ViT encoder, and two-stage fine-tuning with curated detailed captions	Vision-Language alignment / multimodal generation	-	66.2%
[3]	2025	WasteNet and TACO	7 Categories	CLIP	Image resizing and normalization	Multi-class classification	90%	97.18%

3. CLASSIFICATION MODEL TRAINING AND EVALUATION

3.1 The Model

The proposed system is based on the MobileNetV2 architecture, chosen for its lightweight design and efficiency in mobile and low-resource environments. It employs depth-wise separable convolutions which divide convolution into two parts: depth-wise and point-wise operations to decrease parameter count while sustaining system efficiency [27].

The transfer learning approach utilizes the MobileNetV2 model which has been pretrained on ImageNet to enhance performance in situations with restricted data. The existing classification layer has been deleted to create a new head which includes pooling, normalization, dropout, and dense layers. The new layers of the system will undergo training first while keeping the pretrained layers in their current state. The model undergoes fine-tuning after its initial training through the process of unfreezing deeper layers which enables the system to learn the waste classification task. The stable training process uses low learning rates and early stopping and learning rate reduction techniques to achieve better generalization results.

3.2 Dataset

The two publicly available datasets used to train and evaluate the waste classification model were TrashNet [24] and the Garbage Classification dataset [23]. These datasets were combined during processing to improve the diversity and representation of the dataset. The details of each dataset are presented in Table 2 and Table 3.

Table 2. : TrashNet dataset

Category	Number of Images
Cardboard	403
Glass	501
Metal	410
Paper	594
Plastic	482
Trash	137
Total: 2527 images	

Table 3. : Garbage Classification dataset

Category	Number of Images
Cardboard	891
Brown glass	607
Green glass	629
White glass	775
Metal	769
Paper	1,050
Plastic	865
Trash	697
Total: 6283 images	

3.3 Image Augmentation

To improve model generalization and reduce overfitting, data augmentation was applied to the training dataset using the ImageDataGenerator function from Keras. Table 4 summarizes the augmentation experiments conducted in this study.

Table 4. : Summary of Data Augmentation Experiments

Experiment	**Augmentation Techniques**	**Validation Accuracy**	**Test Accuracy**
Exp 1	Rotation ±20°, shift 10%, shear 10%, zoom 20%, brightness [0.8–1.2], horizontal flip	0.9298	0.9155
Exp 2	Rotation ±20°, shift 15%, shear 15%, zoom 20%, brightness [0.8–1.2], horizontal flip	0.9058	0.8902
Exp 3	Rotation ±20°, shift 10%, shear 15%, zoom 20%, brightness [0.8–1.2], horizontal flip	0.9229	0.9139
Exp 4	Rotation ±30°, shift 20%, shear 20%, zoom 20%, brightness [0.8–1.2], horizontal flip	0.9092	0.8767
hline Exp 5	Rotation ±20°, shift 10%, shear 10%, zoom 25%, brightness [0.9–1.1], horizontal flip	0.9195	0.9088
Exp 6	Rotation ±10°, shift 10%, shear 10%, zoom 10%, brightness [0.8–1.2], horizontal flip	0.9195	0.9189
Exp 7	Rotation ±15°, shift 10%, shear 8%, zoom 15%, brightness [0.9–1.1], horizontal flip + noise	0.8921	0.8564
Exp 8	Rotation ±15°, shift 10%, shear 5%, zoom 10%, brightness [0.95–1.05], horizontal flip	0.8887	0.8632
Exp 9	Rotation ±20°, shift 10%, shear 15%, zoom 20%, brightness [0.8–1.2], horizontal flip	0.9269	0.9241
Exp 10	Rotation ±30°, shift 15%, shear 20%, zoom 25%, brightness [0.8–1.2], horizontal + vertical flip, channel shift 20.0, fill mode nearest	0.9425	0.9396
Best Model	**Optimized Augmentation (Exp 10)**	**94.25%**	**93.96%**

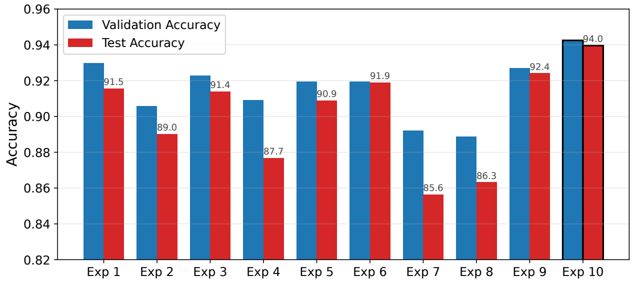


Fig. 1: Validation and test accuracy obtained under different data augmentation configurations.

As illustrated in Figure 1, augmentation intensity had a substantial impact on model performance and generalization. Experiments employing aggressive transformations, such as Exp 4 and Exp 7, produced noticeably lower validation and test accuracies compared with other configurations. In particular, Exp 7 achieved the lowest test accuracy (85.64%), suggesting that noise injection introduced unrealistic patterns that negatively affected feature learning. Similarly, Exp 4 showed reduced performance due to the use of stronger geometric transformations, which may have distorted important visual characteristics of waste materials.

Conversely, moderate augmentation settings generally resulted in better generalization. Exp 9 and Exp 10 achieved the highest performance among all evaluated configurations, with test accuracies of 92.41% and 93.96%, respectively. These findings indicate that carefully balanced augmentation strategies can improve model robustness while preserving the essential characteristics of the original images.

Data augmentation experiments demonstrated that augmentation intensity has a significant impact on model performance. Excessive augmentation, such as increasing image rotation beyond $\pm 20^\circ$ or applying large shift and shear values above 15%, distorted the original image structure and negatively affected classification accuracy. Similarly, applying strong noise augmentation introduced unrealistic visual patterns that reduced the model's ability to learn meaningful features.

In contrast, insufficient augmentation, such as using a very limited brightness range (e.g., [0.95–1.05]) with minimal zoom or shear transformations, did not provide enough variability in the training data, leading to weaker generalization performance.

Therefore, the experimental results suggest that moderate augmentation settings achieve the best balance between dataset diversity and image realism, leading to improved classification accuracy and generalization performance. Both excessive and insufficient augmentation were found to reduce overall model effectiveness.

3.4 Pre-processing

Before training the model, several preprocessing steps were applied to prepare the input data. First, all images were resized to 224×224 pixels to match the input size required by the MobileNetV2 architecture.

Next, pixel values were normalized using the *preprocess_input* function provided by MobileNetV2. This step scales the input images to the same distribution used during ImageNet pretraining, which helps the model effectively utilize the pretrained weights and improves learning performance.

The preprocessing step was applied consistently to the training, validation, and test datasets to ensure uniform input representation.

3.5 Experimental Setup

The MobileNetV2 model was trained using the Adam optimizer with categorical cross-entropy and label smoothing to improve generalization and reduce overfitting. Training used EarlyStopping and ReduceLROnPlateau and ModelCheckpoint as callbacks to maintain stability and perform dynamic learning rate adjustments and save the best-performing model.

A two-phase training strategy was applied, starting with a higher learning rate for feature extraction. The training process moved to a lower rate during fine-tuning which helped to maintain pretrained knowledge and boost model convergence. The model was developed using TensorFlow/Keras on Google Colab Pro which provided GPU support to achieve fast training times and dependable system operation.

3.6 Results and Discussion

3.6.1 Evaluation Metrics

To comprehensively assess the performance of the proposed model, four standard classification metrics were employed: accuracy, precision, recall, and F1-score. These metrics are computed from the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) produced by the classifier on the test set.

Accuracy measures the overall proportion of correctly classified samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision quantifies how many of the samples predicted as a given class actually belong to that class:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall (sensitivity) quantifies how many of the actual samples of a class were correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

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

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Since the task involves six classes, precision, recall, and F1-score are computed for each class in a one-vs-rest manner and summarized using macro and weighted averages, where the macro average treats all classes equally and the weighted average accounts for class support. In addition, a confusion matrix is used to analyze the distribution of misclassifications across categories.

3.6.2 Overall Performance

Table 5 presents the accuracy results of the MobileNetV2 model on the training, validation, and test datasets. The model achieved high accuracy across all phases, indicating effective learning during training and good generalization when evaluated on unseen data.

Table 6 reports the detailed classification report of the model on the test set, including the per-class precision, recall, and F1-score, together with the support of each class. The model maintained an F1-score above 0.88 across all six categories, achieving macro- and weighted-average F1-scores of 0.9357 and 0.9400, respectively.

Table 5. : Accuracy results of the MobileNetV2 model on training, validation, and test datasets

Metric	Accuracy
Training	97.07%
Validation	94.25%
Test	93.96%

Table 6. : Classification report of the model on the test data

Classification	Precision	Recall	F1-Score	Support
Cardboard	0.9556	0.9556	0.9556	90
Glass	0.9792	0.9261	0.9519	203
Metal	0.9000	0.9351	0.9172	77
Paper	0.9714	0.9714	0.9714	105
Plastic	0.8587	0.9080	0.8827	87
Trash	0.9195	0.9524	0.9357	84
Accuracy		0.9396		646
Macro Avg	0.9307	0.9414	0.9357	646
Weighted Avg	0.9412	0.9396	0.9400	646

3.6.3 Per-Class Performance

A closer inspection of the per-class results in Table 6 and Figure 2 reveals clear differences in difficulty across waste categories. Paper and Cardboard achieved the most balanced performance (F1-scores of 0.9714 and 0.9556, respectively), which can be attributed to their distinctive textures and relatively consistent appearance across the dataset.

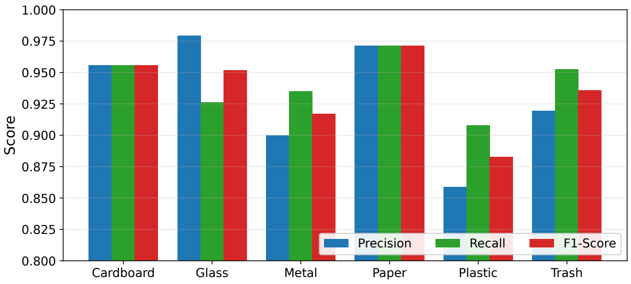
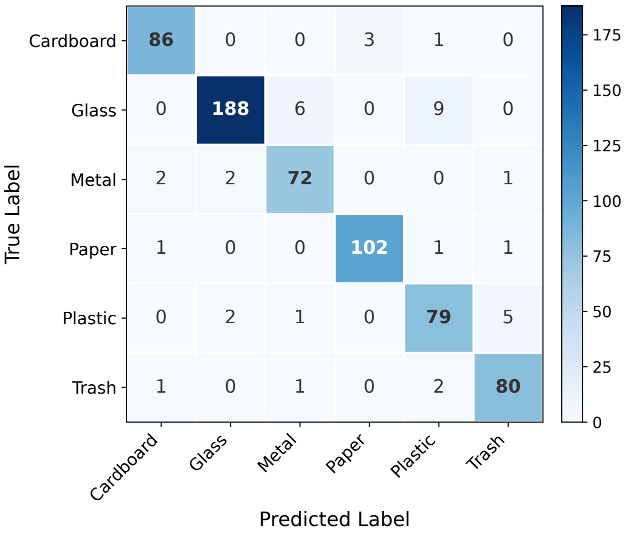


Fig. 2: Per-class precision, recall, and F1-score of the proposed model on the test set.

Plastic was the most challenging category, with the lowest precision (0.8587) and F1-score (0.8827). This behavior is expected, as plastic items exhibit high intra-class variability in shape, color, and transparency, and transparent plastic containers share strong visual similarities with glass. This is further reflected in the Glass category, which achieved the highest precision (0.9792) but a comparatively lower recall (0.9261), suggesting that a portion of glass samples were misclassified into visually similar categories such as plastic. Metal showed a similar pattern in reverse (precision 0.9000, recall 0.9351), indicating that reflective surfaces of other materials were occasionally predicted as metal.

Despite these challenges, the macro-averaged F1-score of 0.9357 demonstrates that the model maintains consistently strong performance across all six categories rather than being biased toward well-represented classes.

3.6.4 Confusion Matrix Analysis



True Label	Cardboard	Glass	Metal	Paper	Plastic	Trash
Cardboard	**86**	0	0	3	1	0
Glass	0	**188**	6	0	9	0
Metal	2	2	**72**	0	0	1
Paper	1	0	0	**102**	1	1
Plastic	0	2	1	0	**79**	5
Trash	1	0	1	0	2	**80**

Fig. 3: Confusion matrix of the proposed model on the test set.

To gain deeper insight into the model's misclassification behavior, the confusion matrix on the test set is shown in Figure 3. The most frequent confusion occurred between Glass and Plastic, with 9 glass samples predicted as plastic and 6 as metal. These errors are consistent with the visual similarity between transparent or reflective materials, which share comparable color distributions and surface characteristics. Likewise, 5 plastic samples were misclassified as trash, which can be explained by the heterogeneous nature of the trash category that includes deformed or mixed-material items. Importantly, the misclassifications are concentrated among visually similar material pairs, while semantically distant categories (e.g., Paper vs. Metal) exhibit almost no confusion. This indicates that the model has learned meaningful material-specific features rather than relying on superficial cues, and that the remaining errors stem primarily from inherent visual ambiguity rather than model limitations.

3.6.5 Training Behavior

The training and validation accuracy and loss curves for both the initial training and fine-tuning phases are shown in Figures 4 and 5. The results show that the model achieved stable learning behavior during training, as the training loss gradually decreased while the accuracy improved across epochs.

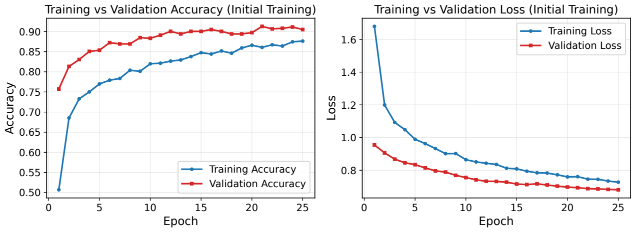


Fig. 4: Training and validation accuracy and loss during initial training

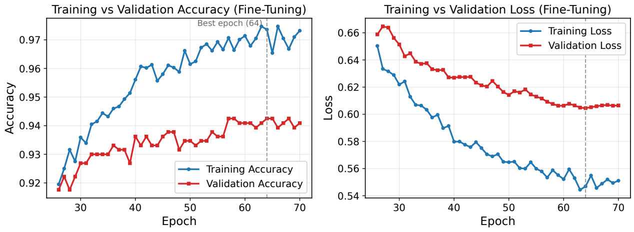


Fig. 5: Training and validation accuracy and loss during fine-tuning

Although a small gap between training and validation performance was observed during fine-tuning, the model still maintained strong generalization and achieved high classification performance on the test dataset.

It can also be observed that the validation loss remains slightly below the training loss during the initial epochs. This behavior is expected, as data augmentation, dropout, and label smoothing are applied only to the training data, making the training task intentionally harder than the validation task. The early-stopping mechanism restored the best model weights from epoch 64, when the validation accuracy reached its maximum of 94.25%.

4. MOBILE APPLICATION DEVELOPMENT

The workflow of the proposed system, *ReLeaf*, is illustrated in Figure 6. This section describes the development of the mobile application and its main components. The following subsections present the tools and platform used, as well as the integration of the large language model (LLM).

4.1 Tools and Platform

The ReLeaf application was developed using a combination of modern tools and platforms to ensure efficient model development, deployment, and user interaction.

For the mobile application, Flutter was used as the primary development framework due to its cross-platform capabilities and ability to create responsive, user-friendly interfaces. The application was developed and tested using Visual Studio Code and Android Studio.

The waste classification model was implemented using TensorFlow and Keras. It was trained and fine-tuned in Google Colab Pro, which provides GPU support to accelerate training and improve performance. Pretrained ImageNet weights were utilized to enhance feature extraction through transfer learning. For deployment on mobile devices, the trained model was converted into TensorFlow Lite (TFLite) format, enabling efficient real-time inference with reduced latency and lower computational requirements. To evaluate real-time performance, the deployed model was tested on a physical mobile device, allowing accurate assessment of inference speed, responsiveness, and overall user experience under real-world conditions.

Firebase was used as the backend platform to manage user authentication and data storage. Firebase Authentication handled secure user login, while Cloud Firestore stored application data, including user information and system interactions.

In addition, a cloud-based Large Language Model (LLM), specifically GPT-4.1-mini, was integrated into the system to provide users with intelligent responses and recycling guidance. This enhances user interaction by allowing natural language communication within the application.

Overall, the combination of these tools and platforms enabled the development of a scalable, efficient, and user-friendly smart waste classification system.

4.2 LLM Integration

To enhance user interaction and provide intelligent recycling guidance, a Large Language Model (LLM) was integrated into the ReLeaf application. The system utilizes the GPT-4.1-mini model through a cloud-based API to enable natural language communication between the user and the application.

The LLM component allows users to ask questions related to waste classification, recycling methods, and proper disposal practices. Based on the user query, the model generates informative and context-aware responses to assist users in making environmentally responsible decisions.

The chatbot interface was integrated into the Flutter application to provide a simple and interactive user experience. User messages are sent securely to the API, and the generated responses are displayed in real time within the application.

The integration of the LLM improves the usability of the system by extending its functionality beyond image classification, enabling users to receive additional guidance and educational support related to recycling and waste management.

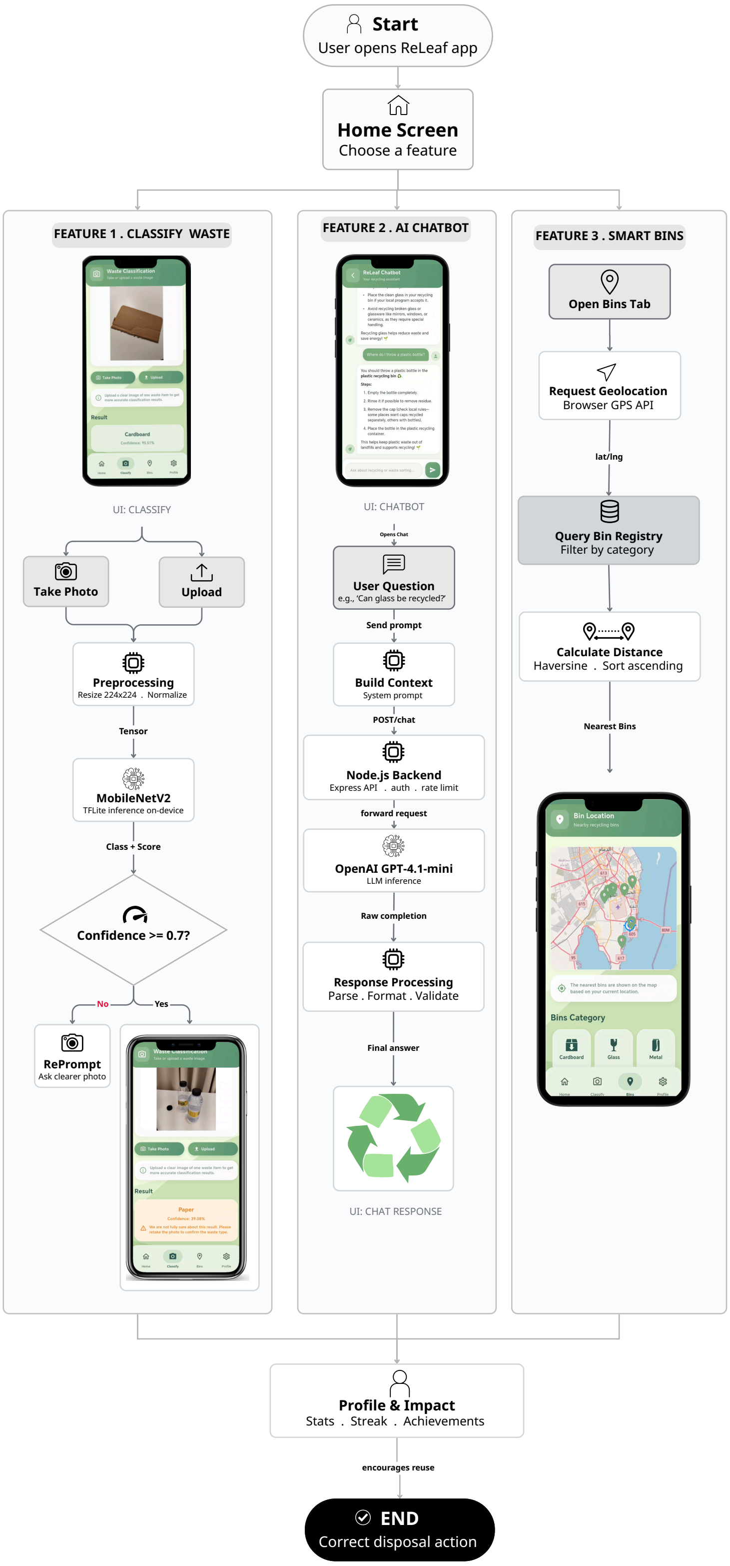


Fig. 6: Workflow of the proposed ReLeaf system integrating image classification and chatbot assistance.

5. CONCLUSION AND FUTURE WORK

This paper presents a smart waste classification approach that integrates deep learning with a large language model. The dataset was constructed by combining and preprocessing multiple public datasets, then split into training, validation, and testing sets. A MobileNetV2 model with transfer learning and fine-tuning was employed for efficient feature extraction, achieving approximately 93.96% test accuracy with minimal confusion between visually similar waste categories.
The trained model was deployed within *ReLeaf*, a cross-platform mobile application built using Flutter and TensorFlow Lite, enabling real-time, on-device waste classification. In addition, a cloud-based LLM (GPT-4.1-mini) was integrated to provide explanations and recycling recommendations, selected for its performance and reliability. Overall, the proposed approach demonstrates an effective end-to-end solution that combines accurate image-based classification with intelligent, conversational guidance to support informed decision-making in waste disposal and promote sustainable recycling practices.

5.1 Future Work

Several directions remain for future work:

—**Extended classification scope:** The model can be extended to support a larger number of waste categories, including hazardous and electronic waste, and to handle images containing multiple objects through object detection techniques such as YOLO.

—**Dataset expansion:** Expanding the training data with locally collected images under diverse real-world conditions (varying lighting, backgrounds, and device cameras) would further improve generalization to real-world deployment scenarios.

—**Enhanced LLM capabilities:** The LLM component can be enhanced with multilingual support, particularly Arabic, and on-device inference to reduce latency and dependence on cloud services.

—**Smart city integration:** Integrating ReLeaf with IoT-enabled smart bins and municipal recycling infrastructure would enable automated, city-scale waste management.

—**User study:** A longitudinal user study could evaluate the application's real-world impact on recycling behavior and user engagement over time.

The source code and implementation details are available at GitHub: `https://github.com/fatemahbarri/releaf_app`

6. REFERENCES

[1] Aulia Afifah, Endah Ratna Arumi, Maimunah Maimunah, Setiya Nugroho, et al. Comparative evaluation of preprocessing methods for mobilenetv1 and v2 in waste classification. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 9(3):194–202, 2025.

[2] Gulzar Ahmad, Fizza Muhammad Aleem, Tahir Alyas, Qaiser Abbas, Waqas Nawaz, Taher M Ghazal, Abdul Aziz, Saira Aleem, Nadia Tabassum, and Aidarus Mohamed Ibrahim. Intelligent waste sorting for urban sustainability using deep learning. *Scientific reports*, 15(1):27078, 2025.

[3] Iman Ameer Ahmad, Ahmed Mudheher Hasan, and Amjad Jaleel Humaidi. Development of a memory-efficient and computationally cost-effective cnn for smart waste classification. *Journal of Engineering Research*, 2025.

[4] Imam Husni Al Amin, Dwiki Vio Setyadarma, and Setyawan Wibisono. Integration of the faster r-cnn algorithm for waste detection in an android application. *Revue d'Intelligence Artificielle*, 37(6):1407, 2023.

[5] Nur Zulaikha Binti Azhar, Noor Hasimah Ibrahim Teo, Raseeda binti Hamzah, Rosniza binti Roslan, and Ruhaila Maskat. A hybrid resnet—mobilenet deep learning model for smart bin waste classification. In *2024 5th International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, pages 356–362. IEEE, 2024.

[6] Oriol Chacón-Albero, Mario Campos-Mocholí, Cédric Marco-Detchart, Vicente Julian, Jaime Andrés Rincon, and Vicent Botti. Ai for sustainable recycling: Efficient model optimization for waste classification systems. *Sensors*, 25(12):3807, 2025.

[7] Tio Dharmawan, Yudha Alif Auliya, Diah Ayu Retnani, Saiful Bukhori, Mohammad Zarkasi, and Imanuel Ataama. Optimizing densenet121 for waste classification using genetic algorithm-based downsampling and data augmentation. *Mathematical Modelling of Engineering Problems*, 12(5), 2025.

[8] Ricy Firnando, Muhammad Ali Buchari, Anna Dwi Marjusalinah, Rahmat Fadli Isnanto, et al. Performance analysis of mobilenetv2 based automatic waste classification using transfer learning. *Jurnal Mandiri IT*, 14(1):257–268, 2025.

[9] Akshat Gaurav, Brij Bhooshan Gupta, Varsha Arya, Razaz Waheeb Attar, Shavi Bansal, Ahmed Alhomoud, and Kwok Tai Chui. Smart waste classification in iot-enabled smart cities using vgg16 and cat swarm optimized random forest. *Plos one*, 20(2):e0316930, 2025.

[10] Hafizatul Hanin Hamzah, Wan Athirah Wan Endut, Azmina Mohd Ariffin, and Nur Atiqah Sia Abdullah. Waste management classification using convolutional neural network. In *International Conference on Innovation & Entrepreneurship in Computing, Engineering & Science Education (InvENT 2024)*, pages 330–340. Atlantis Press, 2024.

[11] Teny Handhayani and Janson Hendryli. Leboh: An android mobile application for waste classification using tensorflow lite. In *Proceedings of SAI Intelligent Systems Conference*, pages 53–67. Springer, 2022.

[12] Teny Handhayani, Ageng Hadi Pawening, and Janson Hendryli. Leboh 2: An android application for solid waste detection. *IAENG Int J Comput Sci*, 50(4):1–9, 2023.

[13] Md Ibnul Hasan, Rabiei Mamat, and Mohammad Aizat Bin Basir. Advanced techniques for waste classification using a heterogeneous and hybrid ensemble of convolutional neural networks and parallel processing. *Journal of Engineering and Science Research*, 8(2):08–18, 2024.

[14] Md Mosarrof Hossen, Molla E Majid, Saad Bin Abul Kashem, Amith Khandakar, Mohammad Nashbat, Azad Ashraf, Mazhar Hasan-Zia, Ali K Ansaruddin Kunju, Saidul Kabir, and Muhammad EH Chowdhury. A reliable and robust deep learning model for effective recyclable waste classification. *IEEE Access*, 12:13809–13821, 2024.

[15] Kai Huang, Huan Lei, Zeyu Jiao, and Zhenyu Zhong. Recycling waste classification using vision transformer on portable device. *Sustainability*, 13(21):11572, 2021.

[16] Amir Mahmud Husein, Baren Baruna Harahap, Tio Fulalo Simatupang, Karunia Syukur Baeha, and Bintang Keitaro

Sinambela. A hybrid vision transformer model for efficient waste classification. *Jurnal Ilmu Komputer dan Informasi*, 18(2):261–275, 2025.

[17] Md Minhazul Islam, SM Mahedy Hasan, Md Rakib Hossain, Md Palash Uddin, and Md Al Mamun. Towards sustainable solutions: Effective waste classification framework via enhanced deep convolutional neural networks. *Plos one*, 20(6):e0324294, 2025.

[18] Daniel Hogan Itam, Ekwueme Chimeme Martin, and Ibiba Taiwo Horsfall. Enhanced convolutional neural network methodology for solid waste classification utilizing data augmentation techniques. *Waste Management Bulletin*, 2(4):184–193, 2024.

[19] Suman Kunwar. Mwaste: A deep learning approach to manage household waste. *arXiv preprint arXiv:2304.14498*, 2023.

[20] Suman Kunwar, Banji Raphael Owabumoye, and Abayomi Simeon Alade. Plastic waste classification using deep learning: insights from the wadaba dataset. *arXiv preprint arXiv:2412.20232*, 2024.

[21] Hiranya Jeet Malla, Milad Bazli, and Mehrdad Arashpour. Enhancing waste recognition with vision-language models: A prompt engineering approach for a scalable solution. *Waste Management*, 204:114939, 2025.

[22] Dishant Mewada, Eoin Martino Grua, Ciaran Eising, Patrick Denny, Pepijn Van de Ven, and Anthony Scanlan. Zero-shot learning for sustainable municipal waste classification. *Recycling*, 10(4):144, 2025.

[23] M. Mohamed. Garbage classification dataset. `https://www.kaggle.com/datasets/mostafaabla/garbage-classification`, 2020. [Accessed 1 February 2026].

[24] F. Ozkefe. Trashnet dataset. `https://www.kaggle.com/datasets/feyzazkefe/trashnet`, 2022. [Accessed 1 February 2026].

[25] Mujalin Polchan, Ampai Pukao, Tapana Cheunban, and Somnuek Sinthupuan. Waste management system using waste classification on mobile application. In *2023 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, pages 229–233. IEEE, 2023.

[26] Wenxuan Qiu, Chenxin Xie, and Jingui Huang. An improved efficientnetv2 for garbage classification. In *International Conference on Intelligent Computing*, pages 79–90. Springer, 2025.

[27] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[28] Faizul Rakib Sayem, Md Sakib Bin Islam, Mansura Naznine, Mohammad Nashbat, Mazhar Hasan-Zia, Ali K Ansaruddin Kunju, Amith Khandakar, Azad Ashraf, Molla Ehsanul Majid, Saad Bin Abul Kashem, et al. Enhancing waste sorting and recycling efficiency: robust deep learning-based approach for classification and detection. *Neural Computing and Applications*, 37(6):4567–4583, 2025.

[29] C Vyas, K Rathod, K Makvana, A Nimavat, and K Kandoriya. Smart waste sorting using deep learning and image augmentation. *African Journal of Biomedical Research*, 27:5777–5785, 2024.

[30] Liying Yong, Le Ma, Dandan Sun, and Liping Du. Application of mobilenetv2 to waste classification. *Plos one*, 18(3):e0282336, 2023.

[31] Dashun Zheng, Rongsheng Wang, Yaofei Duan, Patrick Cheong-Iao Pang, and Tao Tan. Focus-rcnet: a lightweight recyclable waste classification algorithm based on focus and knowledge distillation. *Visual Computing for Industry, Biomedicine, and Art*, 6(1):19, 2023.

[32] Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*, 2023.