Automated Recyclable Waste Detection using Deep Neural Networks

Nagaleela Gangula[1], J S. Sukanya[2], Nagesh Chilamakuru[3], Kummara Lokeshnath[4], Shabana G[5] and Ganesh G[6]
{nagaleelag.cse@srit.ac.in[1], sukanyaprasanna2009@gmail.com[2], cnagesh.cse@gmail.com[3], lokeshnathever@gmail.com[4], gshabana.cse@srit.ac.in[5], ganeshg.cse@srit.ac.in.[6]}

Assistant Professor, Department of CSE, Srinivasa Ramanujan Institute of Technology, Anantapur, Andhra Pradesh, India[1,2,3,4,5,6]

Abstract. Automatically classifying waste into recyclable and disposable categories is crucial for improving waste management and reducing environmental impact. This study introduces a deep learning system to classify household waste using publicly available datasets across various categories. We evaluate the performance of MobileNetV2, EfficientNet-B3, and Vision Transformer (ViT) in terms of accuracy and efficiency for image classification. To improve the model's reliability and prevent overfitting, standard preprocessing and data augmentation techniques are applied. Assessment of the reliability and performance of the system in handling waste from different categories. The findings show that these advanced neural networks can effectively identify recyclable materials and provide real-time solutions for waste sorting. This method can be integrated into smart bins, recycling centers, and mobile systems, supporting more sustainable waste management practices

Keywords: Deep learning, recyclable waste detection, MobileNetV2, EfficientNet-B3, Vision Transformer (ViT), image type, CNN, waste sorting.

1 Introduction

Due to an increase in global waste generation, proper waste management is necessary. Conventional manual sorting of waste is generally inefficient and error prone, resulting in low reliability for the classification process. Automatizing this process can increase precision and productivity for waste sorting, utilizing deep learning and computer vision. Vision-based classification systems have proved to be very successful in recognizing different types of waste using neural networks.

In this paper an intelligent system capable of classifying household waste as recyclable or waste is introduced, using state-of-the-art deep learning. The system minimizes human involvement and saves waste sorting time by learning models on labelled images. Neural networks like MobileNetV2 or EfficientNet-B3 are great contenders, as they are accurate and efficient. Because of its compact architecture, they are suitable for smart bins and mobile deployment, since they are able to extract meaningful features with lower complexity. In this paper, after pre-processing the DNN-based and the CNN-based models are trained using skillfully processed dataset with grayscale conversion and normalization. These procedures streamline the data, improve image quality, and guarantee reliable real-life use.

Although CNNs are good at capturing local features on the image, recent work on ViTs has

ICITSM-Part I 2025, April 28-29, Tiruchengode, India
Copyright © 2025 EAI
DOI 10.4108/eai.28-4-2025.2357782

shown that attention mechanisms used in transformers are able to capture the global visual structures that can be used for better classification of images. This enables them to differentiate more accurately between visually similar waste categories. While ViTs are computationally more expensive, they can work well with medium-sized data when some pre-processing is done. Consequently, we combine CNN and Vision Transformer (ViT) models for computational efficiency and classification accuracy.

The training data for this study is a labeled waste images dataset of several categories. Existing datasets were usually based on a binary classification, and the limited the diversity and complexity of the datasets. Preprocessing methods including normalization, graying, and controlled augmentation are exploited in our study to alleviate distortion and facilitate the enhancement of learning quality. The performance of each classifier is measured for separating different waste categories. The results show that ViT is promising for complex/ambiguous images, while MobileNetV2 and EfficientNet-B3 are applicable as efficient and stable solutions for real world applications. By creating field-deployable, smart waste [sorting] machines, this research helps put a scientific and technical basis behind an automated recycling technology and a mobile-based application.

2 Literature Review

The rapid growth of waste in the world has led to the popularity of deep learning-based automatic detection of recyclable waste methods. It's been shown in recent works that Convolutional Neural Network (CNN) and transformer architectures greatly enhance the classification accuracy of waste materials.

Ahmad et al. [1] demonstrated the potential of deep learning for large-scale waste management with an intelligent waste sorting system for urban sustainability. Similarly, a strong detection and classification model is presented in [2] that utilized state of art CNNs to improve the recycling scheme. In [3], a CNN based recyclable waste recognition model was presented with strong results on real datasets.

For real time solutions, object detection frameworks such as YOLO have been popular choice. A YOLOv12-based detector was presented in [4], which achieved the highest accuracy in dynamic waste classification. In [5], a comparative study of YOLO models was presented, which demonstrated that lightweight instances are able to combine efficiency with accuracy. Optimized CNNs for waste recycling classification were presented in [6] and enabled taking the next big step into image-based waste sorting.

Benchmarking of pretrained deep models has also been explored. A comprehensive comparison of pretrained CNNs for landfill waste classification was reported in [7], while an evolutionary optimization approach for CNN parameters was proposed in [8]. In another study [9], enhancements to ResNet-50 through feature fusion and depth-separable convolutions improved overall accuracy in garbage classification.

Device-level implementations have also emerged, with [10] presenting the design of an intelligent waste classification device, demonstrating the practicality of embedded AI systems. Focused research on specific categories, such as plastic waste, was performed in [11], where a dataset-driven approach achieved high accuracy. In addition, an alternative sensing approach using thermal imaging was proposed in [12], enabling differentiation between metallic and non-metallic e-waste fractions.

Beyond the technical aspects, social and behavioural dimensions of waste classification have been studied. For example, [13] examined how public education influences residents' willingness to participate in waste classification. More recent frameworks have emphasized sustainable and scalable solutions, such as enhanced CNNs with advanced preprocessing [14] and federated deep learning approaches for smart city deployments [15].

Overall, prior works highlight significant advancements in waste classification through deep learning, ranging from CNN optimization and object detection to device-level implementations and socio-technical integration. However, challenges remain in dealing with visually ambiguous waste, limited dataset diversity, and ensuring robustness in real-world conditions, motivating further research in automated recyclable waste detection.

3 Proposed Methodology

This study focuses on the smart classification of household waste images using modern image recognition methods. As described within the initial drift diagram, this system starts off evolved by means of maintaining a set of pix in corporations that prescribe a diffusion of substances which include plastic boxes, glass fragments, steel objects, paper applications, and biomass materials. Each image is resized to a standard resolution, and data augmentation techniques such as rotation and flipping are applied. This increases dataset diversity and improves the model's ability to recognize objects from different angles. The dataset is split into two parts: one for training and the other for validating the model's reliability. Fig. 1 shows the methodology of proposed work.

To gather huge information from visuals, we use slight identification and high-capacity designs which includes MobileNetV2, performance EfficientNet-B3 and vision Transformer. These models are initially trained on a larger set of images and then fine-tuned to classify different waste categories. This approach minimizes the need for new data and accelerates the learning process. Eventually, the system is tested for its performance for its potential, that is suitable for actual time in computerized recycling systems, along with clever disposal units and environmental collection points.

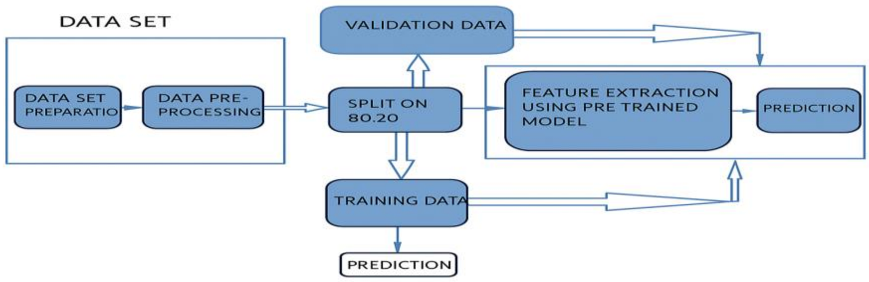

Fig. 1. Methodology of the proposed work

3.1 Dataset Specification

This study uses a comprehensive collection of images representing various categories of household waste, including plastic, metal, glass, paper, cans, and biodegradable materials. The raw images were originally available in different dimensions and formats. To standardize them, all images were resized and reformatted to ensure consistency across the dataset.

Fig. 2. Random images from the Dataset.

Fig. 2 Snap shots are taken from an expansion of lights situations and distinct angles, which improves the compatibility of the machine for the everyday scenes that stumble on waste. For this work, the collection of images is split into two parts: one for training and the other for evaluating its performance. To maintain balance, the dataset was curated so that each category is represented evenly, preventing model bias toward a single waste type. Images were captured under different lighting conditions and angles, which improves the system's adaptability to real-world scenarios. Before training, all images were resized to 224 × 224 pixels and their pixel values normalized to a range between 0 and 1.

To further enhance diversity and robustness, data augmentation techniques such as rotation, flipping, and brightness adjustments were applied. This process increases variability and strengthens the model's ability to generalize when classifying waste materials in diverse environments.

3.2 Data Preprocessing

Before training, several preprocessing steps were applied to prepare the images for effective classification. Each image was resized to 224 × 224 pixels to meet the input requirements of MobileNetV2, EfficientNet-B3, and Vision Transformer (ViT). Pixel values were normalized to the range [0, 1], which improves stability and accelerates training.

To simulate real-world variations, data augmentation methods such as random rotation, horizontal and vertical flipping, and adjustments in brightness and contrast were performed. These augmentations replicate changes in perspective and lighting conditions. In addition, grayscale versions of the images were tested in some experiments to reduce visual complexity and highlight structural features.

These preprocessing steps ensured that the training data was diverse, balanced, and representative of real-world waste classification scenarios.

3.3 Data Splitting

After preprocessing, the complete dataset is divided into two subsets: one for training the models and the other for evaluating their performance on unseen data. Approximately 80% of the images are used for training, while the remaining 20% are reserved for testing.

To ensure fairness and prevent bias, all waste categories are proportionally represented in both subsets. This balanced approach allows the models to learn from a wide variety of examples and ensures reliable performance across different types of waste. By maintaining diversity in both the training and testing sets, the system is better equipped to generalize and deliver accurate results in real-world scenarios.

3.4 Model Architecture

The proposed model architecture is compact and well-suited for efficient visual recognition tasks. The system accepts input images of size 128 × 128 pixels with three color channels. These inputs are processed through multiple 3 × 3 convolutional filters, which capture local visual patterns and structural details. To reduce the spatial dimensions while retaining important features, 2 × 2 pooling operations are applied.

This design ensures that the model extracts meaningful representations of waste images while minimizing computational complexity. The resulting feature maps form the foundation for subsequent classification tasks performed by the selected deep learning models.

3.5 MobileNetV2 Architecture

The MobileNetV2 model is designed to achieve efficient image classification with minimal computational cost. It uses depthwise separable convolutions, which process each feature map individually before combining them, thereby reducing the number of parameters while still capturing important visual information.

In this architecture, the input images are progressively reduced from 64 × 64 pixels to 4 × 4 feature maps through a series of convolutional and pooling operations. The extracted features are then consolidated into a single feature vector. This vector is passed to the final classification layer, which assigns the image to the appropriate waste category.

Figure 3 illustrates the overall architecture of the MobileNetV2 model. This design ensures a balance between accuracy and efficiency, making it well-suited for real-time applications such as smart bins and mobile-based waste classification systems.

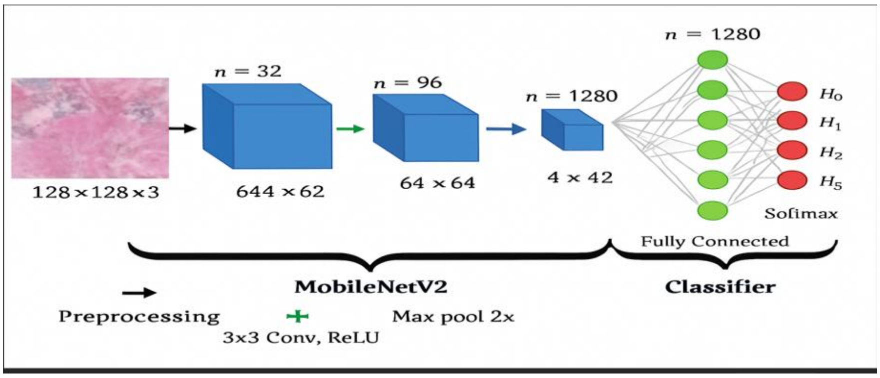

Fig. 3. MobileNetV2 Architecture.

3.6 EfficientNet-B3 Architecture

The EfficientNet-B3 model is employed to enhance waste classification by balancing accuracy and computational efficiency. Input images of size 256 × 256 pixels are passed through multiple convolutional layers, with normalization techniques applied at each stage to stabilize training and optimize performance. The network progressively reduces the spatial dimensions while retaining key visual features, enabling it to capture detailed patterns present in waste images.

During training, dropout layers are introduced to prevent overfitting and improve the model's generalization ability. The extracted features are then flattened and passed through a series of dense layers, where the final classification is determined. This architecture provides a reliable framework for identifying various waste categories, offering strong accuracy even in environments with limited computational resources. The final step of the technique assigns the proper label based totally at the traits of the picture as shown below in Fig 4. This framework is sufficient for conditions with restricted pc energy, supplying a balanced solution that isn't compromised with identity nice.

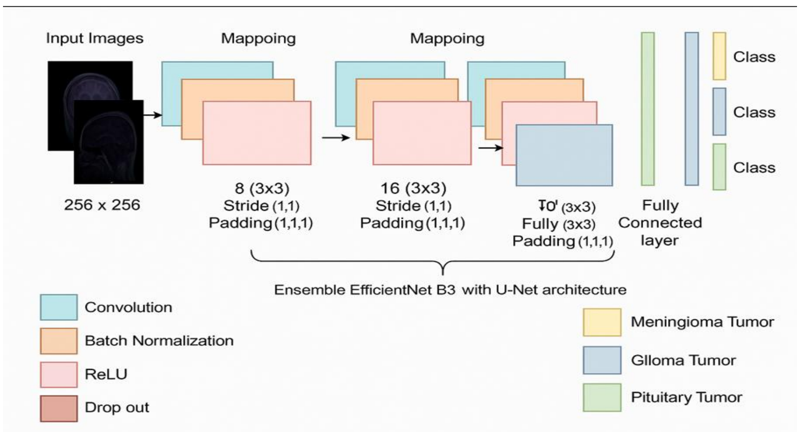

Fig. 4. EfficientNet-B3 architecture.

3.7 Vision Transformer (ViT) Architecture

The photograph of the shape used on this process starts off evolved by using dividing the photo into smaller sections in same quantities. Each image is divided into smaller sections, converted into numerical representations, and marked with additional markers to help the system identify key areas of the image. The sequence of these sections flows thru a gadget, that may estimate the long-distance connections in the entire film. Within these layers, the mechanism focuses on precise regions to fit their importance, enhancing the advent of visible information. Once processed, the very last output passes through a layer of decision to determine the most appropriate category. This strategy is powerful inside the description of images that appear like complex or very excessive to each different. Fig.5 shows the vision transformer architecture.

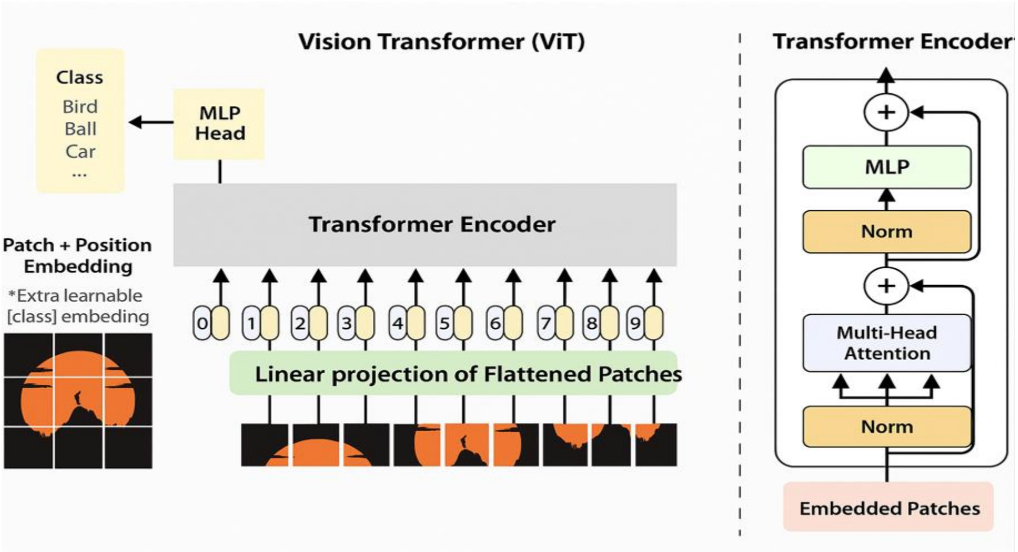

Fig. 5. Vision Transformer (ViT) architecture.

3.8 Performance Metrics

In this study, performance measurements are used to evaluate the accuracy of different types of waste sample classifications. The evaluation is mainly carried out using standard assessment metrics and the confusion matrix. These metrics help determine whether the models are performing reliably and provide insight into areas where classification errors occur.

3.8.1 Evaluation Metrics

Accuracy: Measures the proportion of correctly predicted samples to the total number of predictions, as shown in Equation (1).

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

Precision: Precision focuses on the reliability of positive classifications, measuring the proportion of correctly predicted positive instances out of all instances predicted as positive, as shown in formula (2).

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

Recall (Sensitivity): Recall measures the system's ability to correctly identify actual positive cases, minimizing false negatives, as defined by the recall formula (3).

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

F1-Score: The F1-score combines both precision and recall, providing a balanced measure of the model's classification performance. It is calculated using formula (4).

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

Confusion Matrix: The confusion matrix is a visual tool used to compare the predicted classifications with the actual categories, helping to identify where errors occur. By analyzing the matrix, it is possible to observe how well different classes are recognized and where misclassifications arise, providing insights for improving future performance. Fig 6 shows the confusion matrix.

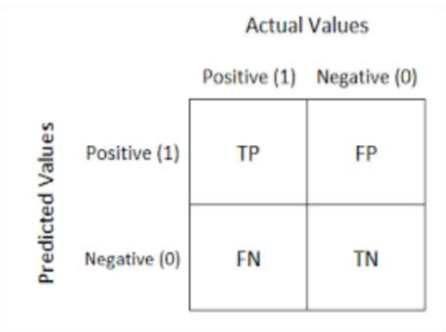

Fig. 6. Confusion Matrix.

4 Results and Discussion

4.1 Evaluation Metrics

The evaluation of the proposed system, which uses deep neural networks for automated waste detection, shows significant progress in accurately identifying and classifying different categories of household waste. Among the models tested, EfficientNet-B3 produced the best results, achieving an overall accuracy of 91%, with precision of 0.91, recall of 0.90, and an F1-score of 0.90.

These values demonstrate its strong capability to classify recyclable and disposable substances consistently and reliably. The Vision Transformer (ViT) closely followed, achieving 90% across all metrics, reflecting its strength in handling visually complex or overlapping waste categories due to its attention-based architecture. Meanwhile, MobileNetV2 delivered stable performance with 87% accuracy, 0.88 precision, 0.87 recall, and 0.87 F1-score, making it a practical option for real-time applications where efficiency and low computational cost are critical.

Although all three models performed effectively, EfficientNet-B3 provided the best overall balance between accuracy and computational efficiency. The results also show that ViT is especially powerful in managing visually challenging classifications. Together, these models highlight the potential for integration into smart bins and mobile applications to support real-world waste management systems.

Table 1. Methods of Algorithm evaluation.

Model	Precision	Recall	F1-Score	Accuracy
MobileNetV2	0.88	0.87	0.87	0.87
EfficientNetB3	0.91	0.9	0.9	0.91
Vision Transformer (ViT)	0.9	0.9	0.9	0.9

4.2 Matrix for confusion

The confusion matrices of the proposed models illustrate how effectively each algorithm classified the different waste categories.

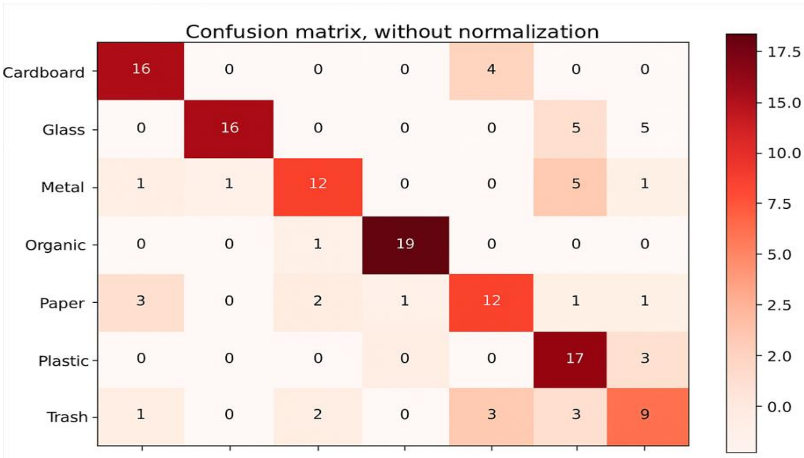

Fig. 7. Confusion Matrix of Transformer Vision (VIT).

Vision Transformer (ViT): As shown in Figure 7, ViT achieved high accuracy in recognizing categories such as carton, glass, and organic waste. However, it showed some confusion between visually similar classes, such as plastic and paper, as well as plastic and metal. Despite these overlaps, ViT demonstrated strong overall performance, particularly in handling complex or ambiguous images.

MobileNetV2: Figure 8 shows that MobileNetV2 correctly identified most categories, especially glass, organic, and carton. However, misclassifications occurred between plastic and metal, as well as occasional errors in the scrap category. While less accurate than EfficientNet-B3 and ViT, MobileNetV2 remains a reliable model for real-time applications due to its efficiency and low computational cost.

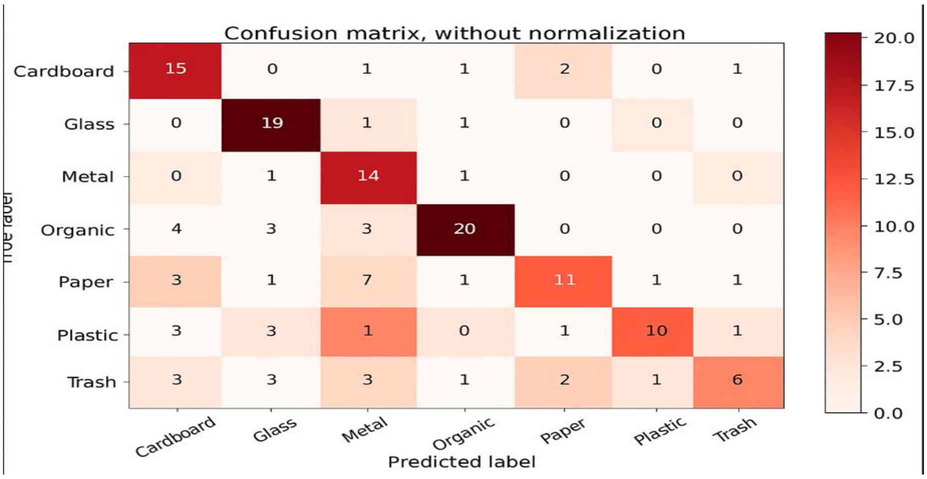

Fig. 8. Confusion Matrix of MobileNetV2.

EfficientNet-B3: Figure 9 highlights that EfficientNet-B3 achieved the best balance across all categories, with minimal misclassifications. It performed strongly in identifying metallic, organic, and paper waste. Although minor confusion was observed between plastic and carton, these errors were less frequent compared to other models. Overall, EfficientNet-B3 demonstrated the highest reliability and robustness, making it the most suitable model for practical deployment in automated waste sorting systems.

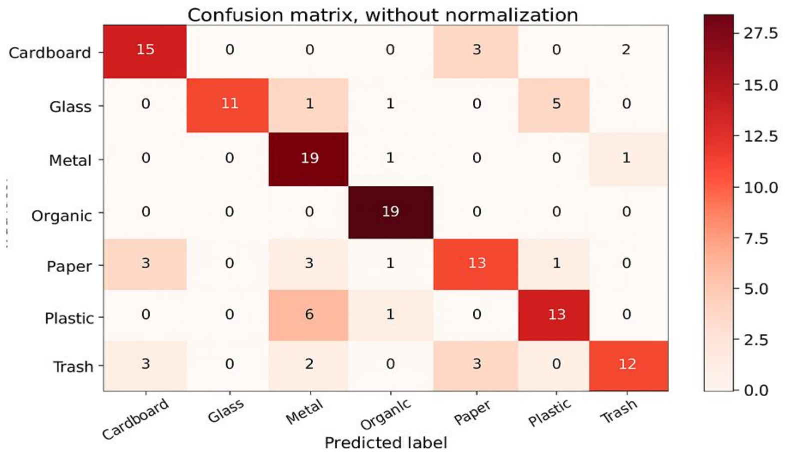

Fig. 9. Confusion Matrix of EfficientNet-B3.

5 Conclusion

This study demonstrates the potential of deep learning for automating the classification of household waste into recyclable and non-recyclable categories. Among the tested models, EfficientNet-B3 achieved the highest performance, with 91% accuracy, precision, recall, and F1-score, making it the most effective choice for real-world applications. The Vision Transformer (ViT) also performed strongly, especially in distinguishing visually complex waste categories, while MobileNetV2 provided efficient results suitable for real-time deployment in resource-constrained environments.

The evaluation metrics and confusion matrices confirm the robustness of these models, with fewer misclassifications and consistent accuracy across different waste categories. These results highlight the promise of deep learning in reducing manual effort, improving classification efficiency, and contributing to sustainable waste management practices.

Future work will focus on expanding the dataset to include additional waste categories, improving classification of visually similar materials, and implementing real-time solutions in smart bins and recycling centers. This research provides a strong foundation for developing intelligent, scalable, and environmentally impactful waste sorting systems.

References

[1] Ahmad, G., Aleem, F. M., Alyas, T., & others. (2025). Intelligent waste sorting for urban sustainability using deep learning. Scientific Reports, 15, 27078. https://doi.org/10.1038/s41598-025-08461-w

[2] Sayem, F. R., Islam, M. S. B., Naznine, M., & others. (2025). Enhancing waste sorting and recycling efficiency: Robust deep learning-based approach for classification and detection. Neural Computing and Applications, 37, 4567–4583. https://doi.org/10.1007/s00521-024-10855-2

[3] Zhang, Q., Zhang, X., Mu, X., Wang, Z., Tian, R., Wang, X., & Liu, X. (2021). Recyclable waste

image recognition based on deep learning. Resources, Conservation and Recycling, 171, 105636. https://doi.org/10.1016/j.resconrec.2021.105636

[4] Dipo, M. H., Farid, F. A., Mahmud, M. S. A., Momtaz, M., Rahman, S., Uddin, J., & Karim, H. A. (2025). Real-time waste detection and classification using YOLOv12-based deep learning model. Digital, 5(2), 19. https://doi.org/10.3390/digital5020019

[5] Riyadi, S., Andriyani, A. D., & Masyhur, A. M. (2024). Classification of recyclable waste using deep learning: A comparison of YOLO models. Revue d'Intelligence Artificielle, 38(4), 1089–1096. https://doi.org/10.18280/ria.380404

[6] Mao, W., Chen, W., Wang, C., & Lin, Y. (2021). Recycling waste classification using optimized convolutional neural network. Resources, Conservation and Recycling, 164, 105132. https://doi.org/10.1016/j.resconrec.2020.105132

[7] Younis, H., & Obaid, M. (2024). Performance comparison of pretrained deep learning models for landfill waste classification. International Journal of Advanced Computer Science and Applications, 15(11). https://doi.org/10.14569/IJACSA.2024.0151166

[8] Shah, M. A., Jan, M., & Ullah, A. (2025). Eco-evolution: Transforming waste management through convolutional neural network parameter evolution with optimizers. Cureus, 17(4), e4505. https://doi.org/10.7759/cureus.4505

[9] Li, L., Wang, R., Zou, M., Guo, F., & Ren, Y. (2025). Enhanced ResNet-50 for garbage classification: Feature fusion and depth-separable convolutions. PLOS ONE, 20(1), e0317999. https://doi.org/10.1371/journal.pone.0317999

[10] Chen, Y.-H., & Chao, C.-H. (2025). Design and implementation of an intelligent waste classification device. Computing and Artificial Intelligence, 3(1), 2331. https://doi.org/10.59400/cai2331

[11] Kunwar, S., Owabumoye, B., & Alade, A. (2024). Plastic waste classification using deep learning: Insights from the WaDaBa dataset. https://doi.org/10.48550/arXiv.2412.20232

[12] Gundupalli, S. P., Hait, S., & Thakur, A. (2018). Classification of metallic and non-metallic fractions of e-waste using thermal imaging-based technique. Process Safety and Environmental Protection, 118, 32–39. https://doi.org/10.1016/j.psep.2018.06.022

[13] Liu, X., Wang, Z., Li, W., Li, G., & Zhang, Y. (2019). Mechanisms of public education influencing waste classification willingness of urban residents. Resources, Conservation and Recycling, 149, 381–390. https://doi.org/10.1016/j.resconrec.2019.06.001

[14] Islam, M. M., Hasan, S. M. M., Hossain, M. R., Uddin, M. P., & Mamun, M. A. (2025). Towards sustainable solutions: Effective waste classification framework via enhanced deep convolutional neural networks. PLOS ONE, 20(6), e0324294. https://doi.org/10.1371/journal.pone.0324294

[15] Khan, H. A., Naqvi, S. S., Alharbi, A. A. K., & Alotaibi, S. (2024). Enhancing trash classification in smart cities using federated deep learning. Scientific Reports, 14, 11816. https://doi.org/10.1038/s41598-024-62003-4