Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.12 n.57, p. 1110-1119. January/February, 2026.
DOI: https://doi.org/10.5935/jetia. v12i57.3270

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

AN APPROACH TO AUTOMATION IN SMART WASTE CLASSIFICATION BASED ON MACHINE LEARNING FOR EMBEDDED DEVICES

Pham Trung Kien[1]

[1]Vietnam Institute of Tropical Technology & Environmental Protection (VITTEP), Ho Chi Minh City, Vietnam.

[1]http://orcid.org/0009-0009-9872-9572

Email: phamtrungkien999@gmail.com

ARTICLE INFO

Article History
Received: January 9, 2025
Revised: January 20, 2026
Accepted: January 30, 2026
Published: February 28, 2026

Keywords:
Waste classification,
Embedded devices,
Smart trash bins,
Machine learning,
Raspberry Pi

ABSTRACT

This study aims to develop automated smart waste classification systems using machine learning deployed on resource-constrained hardware, such as Raspberry Pi, with the goal of integrating these models into the initial stage of waste sorting—that is, the trash bin. This study proposes two models: ConvNeXt and an improved CNN model, which is referred to as PrQ-CNN. The improved PrQ-CNN model utilizes pruning and quantization techniques to reduce its size. Both models were trained on a dataset comprising two classes: Non-recyclable and Recyclable. The data were split into a training set and a testing set with a 90:10 ratio. The models were integrated into a Raspberry Pi embedded computer to compare metrics such as model size, inference time, accuracy, missed detection rate (MDR), and false discovery rate (FDR). Experimental results show that the PrQ-CNN model achieved an accuracy of 92%, while its size was 12 times smaller than that of the original model before pruning and quantization. The ConvNeXt model achieved an accuracy of 98%. The results from the improved PrQ-CNN and ConvNeXt models were compared with VGG16, ResNet, EfficientNet, and MobileNet.

I. INTRODUCTION

Waste pollution presents a significant challenge, particularly for developing tropical cities with high population density [1]. Smart cities around the world are currently focusing on solid waste management and treatment [2],[3]. Among these, solid waste sorting and recycling take the lead as major concerns [2]. According to data from the World Bank, global waste volume is projected to increase sharply, reaching 3.4 billion tons by 2050 [4], [5]. Globally, over 43% of solid waste is not properly managed through open burning, illegal dumping, and unregulated landfills [4], [5]. Additionally, the United Nations Environment Programme (UNEP) estimates that manual waste handling causes 15 – 20% of workplace injuries in the recycling sector [4], [5].

Landfilling also produces a large amount of methane, a potent greenhouse gas that contributes significantly to global warming [6]. Methane has a significantly higher heat-trapping potential than carbon dioxide, thereby exerting a stronger impact on global climate change [6]. In summary, if solid waste is not effectively managed, sorted, and recycled, it will lead to increased expenses, time consumption, and can cause waste pollution complications, resulting in consequences such as environmental pollution, impacts on public health, as well as worldwide problems like global warming and climate change. The objective of this study is to build compact machine learning models that can be deployed on resource-constrained hardware, or embedded computers like Raspberry Pi, and other devices, for automated solid waste classification into Recyclable and Non-recyclable types.

This study focuses on optimization, deployment, and evaluation of existing models for embedded waste classification applications. The study also aims to compare these models based on parameters such as performance and size. These hardware platforms can be easily integrated to build smart trash bins that can be deployed in public areas, helping to reduce production costs while introducing artificial intelligence technology into the very first step of waste classification when people discard their waste. Specifically, the waste treatment system model targeted by this study involves building trash bins with processors based on resource-constrained hardware platforms equipped with machine learning models.

Whenever people dispose of waste into these bins, the bins will use artificial intelligence technology to automatically classify the waste as Recyclable or Non-recyclable and separate it into the corresponding section for easier processing and recycling. Accordingly, the main research question addressed in this study is: how can machine learning-based waste classification models be optimized and evaluated to achieve an effective balance between accuracy, model size, and inference speed for deployment on resource-constrained embedded devices.
Contributions

(i) Developing machine learning models for solid waste classification, including a ConvNeXt model on the Kaggle Waste Classification dataset, achieving a high accuracy of 98%.
(ii) Proposing an improved Convolutional Neural Network (CNN) model utilizing pruning and quantization techniques to reduce model size by 12 times compared to the baseline, while maintaining accuracy, making it suitable for resource-constrained hardware and embedded devices.
(iii) Evaluating the models' performance based on accuracy, confusion matrix, and other performance metrics.
(iv) Providing an empirical, deployment-oriented comparison of multiple deep learning models on embedded hardware.

II. RELATED WORKS

According to [7], proposed developing a smart trash bin system capable of real-time waste level monitoring and efficient collection using YOLOv8 and YOLOv5 deep learning models to significantly reduce the number of parameters and processes, thereby lowering hardware requirements. Iman Ranjbar et al. proposed building a large and specialized dataset of plastic waste images, including four types of plastics commonly used in construction projects: ABS, HDPE, PS, and PVC [8]. They leveraged transfer learning with models trained and evaluated on advanced CNNs and Vision Transformers, including ResNet, ResNeXt, RegNet, and Swin Transformer, with Swin Transformer yielding the highest accuracy of 93.28% [8]. Mirna Castro Bello et al. evaluated and compared the effectiveness of four CNN models for waste classification using a custom dataset of 1883 Non-recyclable and Recyclable waste images on a Raspberry Pi 4 Model B with an HQ camera, achieving an average CPU utilization of 86% and an inference time of 1900 ms [9]. Proposed a hybrid method that combines instance segmentation deep learning models with morphological machine learning models to automate the classification and evaluation of decoration and renovation waste on a labeled dataset comprising 53,000 images [10].

The proposed method demonstrated remarkable accuracy, with a low relative error of 2.8% in the overall evaluation between models' predictions and manual sorting [10]. Evaluated the performance of AI models such as Mask R-CNN and YOLOv8, in enhancing plastic waste classification [11]. The models were assessed for accuracy, mean average precision (mAP), precision, recall, F1 score, and inference time, with hyperparameter tuning performed through grid search [11]. Results showed that Mask RCNN achieved an accuracy of 0.912 and mAP of 0.911, performing better than YOLOv8 in tasks requiring detailed segmentation, despite a longer inference time of 200–350 ms [11]. Conversely, YOLOv8, with an accuracy of 0.867 and mAP of 0.922, excelled in real-time scenarios due to its shorter inference time of 80–160 ms [11]. Presented the WasteIQNet model, a hierarchy-aware deep hybrid model, for detailed waste classification [12]. The proposed framework was trained and evaluated on the WEDR dataset, a curated collection of 360,000 images reflecting actual waste conditions across major landfill areas of Delhi [12]. Numerous other studies, such as those in [13-17], are also noteworthy.

These recent and in-depth studies demonstrate researchers' significant interest in the field of waste management and treatment, as well as practical use of new technologies like AI and ML in these areas. Based on the reviewed literature, it is evident that most recent studies primarily concentrate on either (i) multi-class waste classification using large-scale datasets or (ii) achieving high classification accuracy through deep learning models without imposing strict constraints on computational resources. In contrast, only a limited number of studies have systematically examined the trade-offs among model accuracy, model size, and inference speed when deploying waste classification models on resource-constrained embedded platforms, such as the Raspberry Pi. Furthermore, comprehensive comparative evaluations incorporating embedded-oriented performance metrics, including missed detection rate (MDR), false discovery rate (FDR), and on-device inference time, remain scarce in the existing literature.

III. MODELS AND METHODS

III.1 IDENTIFIED RESEARCH GAPS AND CHALLENGES

Based on the literature review, several research gaps can be identified. First, many existing studies prioritize accuracy while overlooking deployment feasibility on embedded hardware with limited memory and computational power. Second, comparisons across models are often conducted on different datasets, making direct evaluation difficult. Third, few studies report embedded-oriented performance indicators such as MDR, FDR, and real inference latency on low-power devices. These gaps motivate this study to focus on practical, deployment-oriented evaluations, with particular emphasis on the implementation of machine learning models for at-source waste classification on resource-constrained embedded systems.

III.2 MODELS AND METHODS

This study focuses on building a solid waste management and treatment system where waste sorting is performed from the very first step, as residents dispose of waste into bins. The trash bin will automatically classify the waste as Recyclable or Non-recyclable using a machine learning model. For such a system, one of the challenges is the limitations of hardware platforms for end-devices. In addition to high accuracy, the model must be sufficiently lightweight and optimized to be seamlessly integrated into these devices. Concretely, the objectives of this study include two main goals:

(i) Building machine learning models for Non-recyclable and Recyclable waste classification that can be applied in the design and manufacturing of smart trash bins for waste management and treatment systems.

(ii) Discussing and proposing the best model that can be applied on resource-constrained hardware platforms for integration into end-user devices, specifically smart trash bins

III.3 SMART TRASH BIN AND DATASET DESCRIPTION

The dataset used in this study is the Waste Classification data from Kaggle. This dataset comprises two components: Training and testing set, split at a 90:10 ratio. Each component consists of two classes: O (Other: Non-recyclable) and R (Recyclable). Figure 1(a) illustrates the structure of the smart waste bin, which employs an AI-based model to perform source-level waste classification. When waste is deposited into the bin, the AI system classifies it into Recyclable and Non-recyclable categories and subsequently controls the conveyor motor to direct the waste into the corresponding compartment. Figure 1(b) presents sample images of the two waste classes, namely Non-recyclable waste and Recyclable waste, contained in the dataset.

(a) Smart trash bin. (b) Images in the dataset.

Figure 1: Smart trash bin structure and sample images from the dataset.

Source: Authors, (2026).

To ensure fair comparison, all models were trained and evaluated using the same dataset split, input resolution (224×224), and evaluation metrics. In this study, data augmentation techniques were not applied during the training process. This design choice was made to ensure that the performance evaluation reflects the model capability under realistic deployment conditions, where waste images captured at the source often exhibit limited variability and are not artificially enhanced. The original class distribution of the dataset was preserved to simulate real-world waste generation scenarios, in which recyclable and non-recyclable waste categories are inherently imbalanced.

Maintaining this natural imbalance allows for a more faithful assessment of the model's robustness and practical effectiveness, particularly in minimizing misclassification risks under real deployment conditions. Additionally, cross-validation was not conducted in this study. Instead, a fixed train–validation–test split was adopted to reduce computational overhead and experimental complexity, which is consistent with the constraints of resource-limited embedded platforms. This approach also aligns with the study's primary objective of deployment-oriented evaluation, where repeatable inference performance, model size, and execution time on embedded hardware are prioritized over exhaustive statistical validation. The hardware specifications used in this study are presented in Table 1.

Table 1: Technical specifications of hardware.

No	Hardware	Technical specifications
1	Raspberry Pi 5	- CPU: ARM Cortex-A76 (ARM v8) 64-bit - Clock speed: 4 x 2.4 GHz - GPU: VideoCore VII (800 MHz) - RAM: 8 GB LPDDR4X (4267 MHz) - SD card: microSD
2	Kaggle GPU T4x2	- Architecture: Turing - CUDA Cores: 2,560 - Tensor Cores: 320 - GPU Memory: 16 GB GDDR6 - Memory Bandwidth: 320 GB/s

Source: Authors, (2026).

III.4 MODEL ARCHITECTURES

III.4.1 PrQ-CNN Model

CNN model: The Convolutional Neural Network model is a type of deep learning algorithm based on multi-layer neural networks that can identify, recognize, and classify objects, as well as detect and segment objects in given images [18].

CNNs typically consist of four types of layers: Convolutional, Pooling, Activation Function, and Fully Connected. The convolutional layer plays a crucial role in the overall CNN structure. It is a set of filters—or kernels—applied to the data. The width, height, and weights of each kernel are used to extract features from the input data. Weight values in the first kernel are randomly assigned but gradually become more informed by the training data. The pooling layer is used to reduce the dimensionality of feature maps while retaining the most important data [18]. A filter applies a pooling operation (max, min avg) by sliding across the input data in the pooling layer. Max pooling is most commonly used. A non-linear layer follows the convolution. Non-linearity allows the generated output to be altered or terminated. This layer is used to constrain or saturate the output [18].

Every type of activation function in every type of neural network serves the essential function of mapping inputs to outputs. The input value is calculated by summing the weighted inputs of the neuron and its bias (if any). This indicates that the activation function determines whether to activate a neuron in response to a given input by producing an appropriate output. Fully Connected (FC) layers, also known as dense layers, are used in neural networks, especially in deep learning. They are a type of neural network layer where every neuron in the layer is connected to every neuron in the preceding and succeeding layers. FC layers are typically deployed at the end of a neural network to perform classification or prediction tasks based on extracted specific features, helping the network learn complex associations between features [18].

Pruning: Pruning is the process of removing unnecessary or less important parts of a model to make it more optimized and faster without significantly reducing its accuracy or performance [19]. Pruning can be categorized into structured, unstructured, and semi-structured pruning. Structured pruning removes filters, channels, or layers to create structured sparsity patterns. In unstructured pruning, individual weight values without structural constraints are considered for removal. Semi-structured pruning acts as a hybrid approach combining aspects of both structured and unstructured pruning [19].

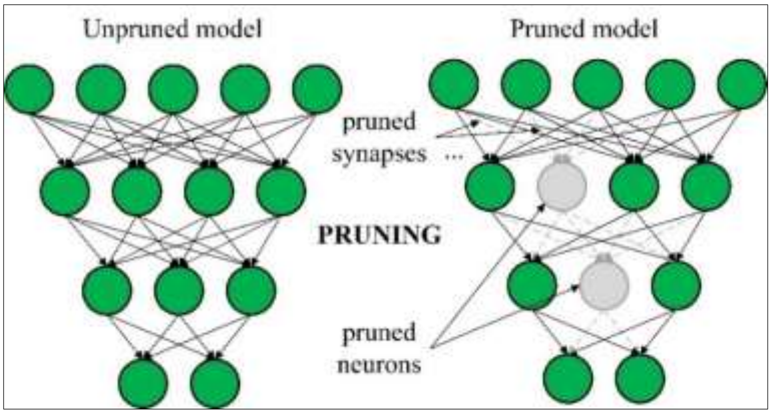

Figure 2: Pruning technique.
Source: Authors, (2026).

Quantization: Quantization is a compression method that converts high-precision representations (32-bit floating point) into lower-bit representations (int8 fixed-point) [20]. Quantization can reduce the CNN model's size, memory footprint, and power consumption, while also improving inference time. This is one of the deep learning model optimization techniques, particularly suitable for embedded systems, mobile devices, or applications requiring high performance and energy efficiency. Quantization techniques used for CNNs are broadly classified as: Post-Training Quantization and Quantization-Aware Training. CNN parameters need to be adjusted to maintain accurate performance after quantization. The process of model retraining to account for quantization is called Quantization-Aware Training, or it can be done without retraining through Post-Training Quantization [20].

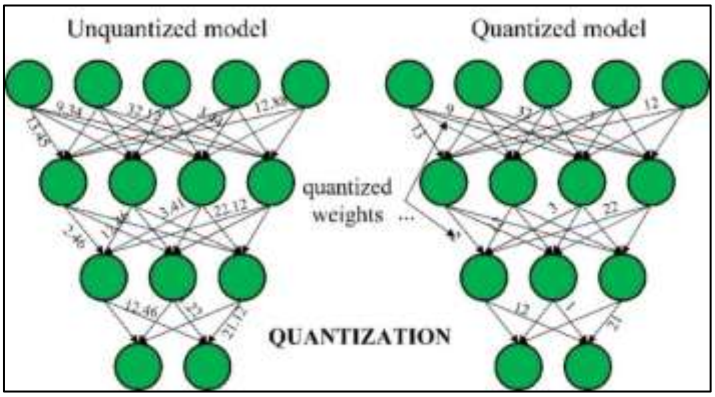

Figure 3: Quantization technique.
Source: Authors, (2026).

In this study, unstructured magnitude-based weight pruning was applied to the convolutional layers, in which individual weights with small absolute values were set to zero. This was followed by post-training dynamic range quantization, where the model weights were quantized to 8-bit integers while activations remained in floating-point precision. The pruning threshold was empirically selected to ensure that the accuracy degradation remained below 2%. Finally, the overall architecture of the PrQ-CNN is shown in Figure 4.

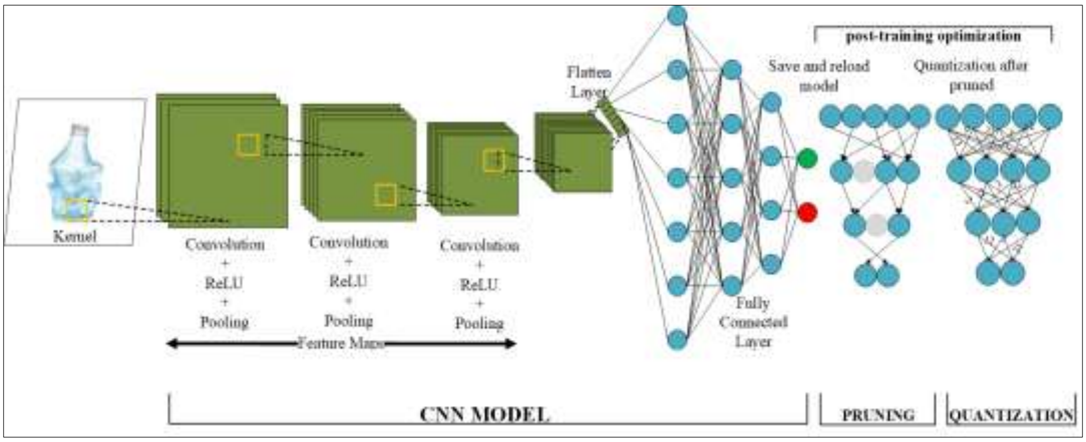

Figure 4: Proposed PrQ-CNN network architecture.
Source: Authors, (2026).

III.4.2 ConvNeXt Model

ConvNeXt is a deep neural network architecture designed to enhance the performance of image processing models, particularly in image recognition and classification [21].

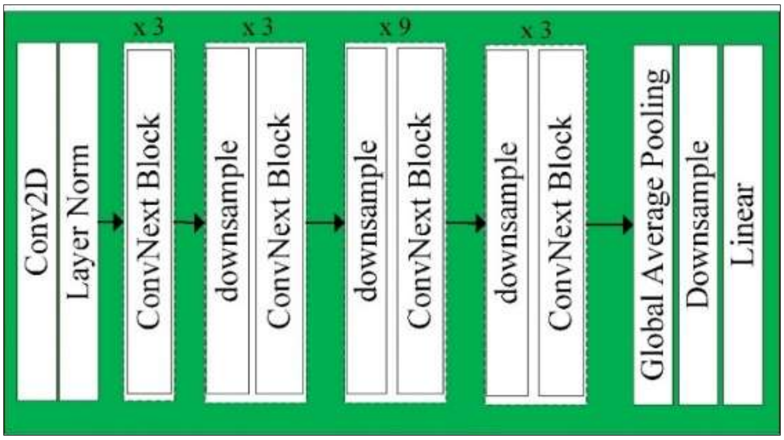

Figure 5: Architecture of the ConvNeXt (tiny) model used in this study.
Source: Authors, (2026).

It was introduced by the Facebook AI Research (FAIR) team in 2022 as a combination of features from traditional CNN models like ResNet and principles of Transformer architecture, aiming to optimize image processing capabilities [21].

IV. RESULTS AND DISCUSSION

IV.1 EVALUATION OF WASTE CLASSIFICATION MODEL RESULTS

To evaluate the performance of the models, this paper calculates and analyzes specific metrics as detailed below:

Confusion Matrix:

A confusion matrix is a specific tabular layout that allows for the visualization of an algorithm's performance. Each row of the matrix represents instances in an actual class, while each column represents instances in a predicted class, or vice versa.

Table 2: Confusion Matrix.

		Predicted Class	
		Yes	No
Actual Class	Yes	TP(a)	FN(b)
	No	FP(c)	TN(d)

Source: Authors, (2026).

Where:
(*) a: TP (true positive) – This occurs when both the actual class and the predicted class of the data point are True.
(*) b: FN (false negative) – This occurs when the actual class of the data point is True, but the predicted class is False (for example, when a pregnancy test indicates that a woman is not pregnant, but she actually is).
(*) c: FP (false positive) – This occurs when the actual class of the data point is False, but the predicted class is True (for example, when a pregnancy test indicates that a woman is pregnant while she is not).
(*) d: TN (true negative) – This occurs when both the actual class and the predicted class of the data point are False.

Accuracy:

Accuracy in classification is the most common evaluation metric. It is defined as the ratio of the number of correct predictions made by the model to the total number of predictions.

Accuracy

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} = \frac{a + d}{a + b + c + d} \tag{1}$$

Error Rate:

The Error Rate of the model is calculated as follows:

Error Rate

$$ErrorRate = 1 - Accuracy \tag{2}$$

Precision:

Precision is defined as the number of true positives divided by the sum of true positives and false positives. It indicates the proportion of positive predictions that are actual positives, in relation to the total number of predicted positives.

Precision

$$Precision = \frac{TP}{TP + FP} = \frac{a}{a + c} \tag{3}$$

Recall/Sensitivity:

Recall is defined as the number of true positives divided by the sum of true positives and false negatives. It indicates the proportion of positive predictions that are actual positives, relative to the total number of positives in the original dataset.

Recall/Sensitivity

$$Recall = \frac{TP}{TP + FN} = \frac{a}{a + b} \tag{4}$$

F1 Score:

The F1 score combines precision and recall for a specific positive class. It reflects the balance between precision and recall, particularly in cases of uneven class distribution. The F1 score achieves its highest value of 1 and its lowest value of 0.

F1 Score

$$F1Score = \frac{2 * Recall * Precision}{Recall + Precision} = \frac{2a}{2a + b + c} \tag{5}$$

MDR (Missed Detection Rate):

MDR is the percentage of fields that are actually positive but are missed (not detected).

MDR

$$MDR = \frac{FN}{TP + FN} = \frac{b}{a + b} \tag{6}$$

FDR (False Discovery Rate):

FDR is the proportion of false positives (incorrect positive detections) among all positive detections.

FDR

$$FDR = \frac{FP}{TP + FP} = \frac{c}{a + c} \tag{7}$$

These models were trained on the Kaggle platform using a T4x2 GPU. The results will be evaluated by implementing the testing set. The improved CNN model (PrQ-CNN) and ConvNeXt will be evaluated alongside VGG16, MobileNet, ResNet, and EfficientNet using performance parameters such as model size (MB), accuracy (%), average inference time per sample (ms), confusion matrix, and other parameters mentioned above on Kaggle GPU and Raspberry Pi. Specifically, the prediction results of the models are listed in Tables 3 to 8.

Table 3: Results of the PrQ-CNN prediction model.

	Precision	**Recall**	**F1-score**	**Support**
Non-recyclable	0.90	0.96	0.93	1401
Recyclable	0.94	0.87	0.91	1112
Accuracy			0.92	2513
Macro avg	0.92	0.91	0.92	2513
Weighted avg	0.92	0.92	0.92	2513

Source: Authors, (2026).

Table 4: Results of the ConvNeXt prediction model.

	Precision	**Recall**	**F1-score**	**Support**
Non-recyclable	0.97	0.99	0.98	1401
Recyclable	0.99	0.96	0.97	1112
Accuracy			0.98	2513
Macro avg	0.98	0.97	0.98	2513
Weighted avg	0.98	0.98	0.98	2513

Source: Authors, (2026).

Table 5: Results of the VGG16 prediction model.

	Precision	**Recall**	**F1-score**	**Support**
Non-recyclable	0.86	0.96	0.91	1401
Recyclable	0.94	0.80	0.87	1112
Accuracy			0.89	2513
Macro avg	0.90	0.88	0.89	2513
Weighted avg	0.90	0.89	0.89	2513

Source: Authors, (2026).

Table 6: Results of the MobileNet prediction model.

	Precision	**Recall**	**F1-score**	**Support**
Non-recyclable	0.85	0.97	0.90	1401
Recyclable	0.94	0.87	0.86	1112
Accuracy			0.88	2513
Macro avg	0.90	0.87	0.88	2513
Weighted avg	0.89	0.88	0.88	2513

Source: Authors, (2026).

Table 7: Results of the ResNet prediction model.

	Precision	**Recall**	**F1-score**	**Support**
Non-recyclable	0.88	0.98	0.92	1401
Recyclable	0.97	0.83	0.89	1112
Accuracy			0.91	2513
Macro avg	0.92	0.90	0.91	2513
Weighted avg	0.92	0.91	0.91	2513

Source: Authors, (2026).

Table 8: Results of the EfficientNet prediction model.

	Precision	**Recall**	**F1-score**	**Support**
Non-recyclable	0.88	0.98	0.92	1401
Recyclable	0.96	0.83	0.89	1112
Accuracy			0.91	2513
Macro avg	0.92	0.90	0.91	2513
Weighted avg	0.92	0.91	0.91	2513

Source: Authors, (2026).

The precision scores for both classes in the ConvNeXt model are the highest (0.97 and 0.99). In contrast, the other models exhibit significant discrepancies between the two classes, with the Non-recyclable class generally performing worse than the Recyclable class. Overall, the ConvNeXt model still demonstrated the best performance, as indicated by its F1-Score, reflecting good balance across both classes. This is followed by PrQ-CNN, which achieved an accuracy of 92%. The confusion matrices of the models are shown in Figure 6.

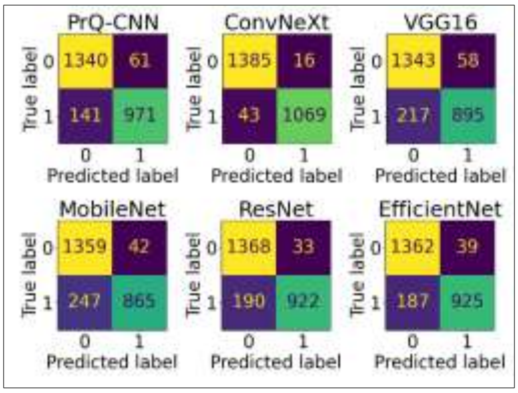

Figure 6: Confusion matrix of the models.

Source: Authors, (2026).

The ConvNeXt model exhibited the most exceptional performance with high accuracy, low error rate, and satisfactory positive class detection. PrQ-CNN followed with strong accuracy. ResNet and EfficientNet demonstrated comparable performance, both commendable but not superior to the ConvNeXt and PrQ-CNN models. The VGG16 and MobileNet models showed lower performance compared to the other models.

IV.2 DISCUSSION

The PrQ-CNN model is relatively small in size (22 MB) and fast, making it suitable for applications requiring quick processing and limited resources, such as embedded computers (92% accuracy). The ConvNeXt model achieved the highest accuracy (98%) but demands more resources than PrQ-CNN. It is ideal for applications where high accuracy is a priority. With an inference time of 234.3 ms/sample on Raspberry Pi, the ConvNeXt model still performs effectively when integrated into embedded devices. In general, the VGG16 and EfficientNet models are larger and slower compared to the other models. MobileNet is the lightest model (9.6 MB), making it suitable for real-time applications or devices with limited resources, though it sacrifices some accuracy, achieving 88% compared to other models. A comparison chart of the models based on six specific parameters is shown in Figure 8.

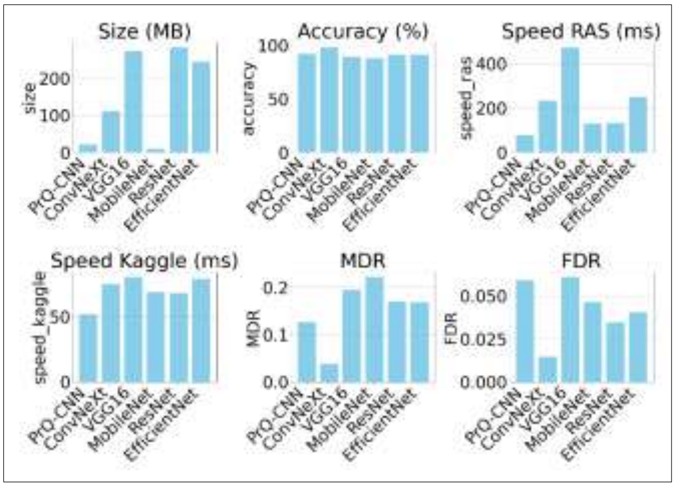

Figure 7: Comparison of models' parameters.
Source: Authors, (2026).

Tables 9 and 10 evaluate the performance of the models.

Table 9: Comparison table of best-performing models.

No	Parameter	Value	Model
1	Size	9.6 MB	MobileNet
2	Accuracy	98.0%	ConvNeXt
3	Speed on Raspberry Pi	76.3 ms	PrQ-CNN
4	Speed on Kaggle	52.0 ms	PrQ-CNN
5	MDR	0.0387	ConvNeXt
6	FDR	0.0147	ConvNeXt

Source: Authors, (2026).

Table 10: Comparison table of worst-performing models.

No	Parameter	Value	Model
1	Size	283.4 MB	ResNet
2	Accuracy	88.0%	MobileNet
3	Speed on Raspberry Pi	472.2 ms	VGG16
4	Speed on Kaggle	80.4 ms	VGG16
5	MDR	0.2221	MobileNet
6	FDR	0.0609	VGG16

Source: Authors, (2026).

Based on the two comparison tables above, several conclusions can be drawn: The ConvNeXt model achieves the best accuracy, while PrQ-CNN remains the leading model in terms of inference time on embedded computers. If accuracy is prioritized, ConvNeXt and PrQ-CNN are top choices. If a lightweight and fast-processing model for resource-constrained devices is required, PrQ-CNN is the most suitable. In general, PrQ-CNN provides a good balance between accuracy, size, and inference time for embedded devices. Table 11 compares the models proposed in this study with those reported in previous studies. The accuracy is evaluated on the same dataset as well as on different datasets. Overall, the ConvNeXt and PrQ-CNN models in this study demonstrated strong performance compared to the studies listed in Table 11. Both models exhibited superior accuracy, with PrQ-CNN leveraging pruning and quantization techniques to reduce the model size by approximately 12 times compared to the baseline unoptimized CNN model, making it suitable for embedded devices with limited hardware resources.

Table 11: Comparison with other models.

No	Study	Model	Dataset	Accuracy
1	This study	ConvNeXt	Kaggle	98.0%
2	This study	PrQ-CNN	Kaggle	92.0%
3	[22]	MLH-CNN	TrashNet	92.6%
4	[23]	EfficientNet	Custom Kaggle	85.0%
5	[24]	CNN	Same dataset	80.88%
6	[25]	CNN	Same dataset	91.32%
7	[26]	YOLOv7-tiny	Various sources	86.8%

Source: Authors, (2026).

V. FUTURE RESEARCH DIRECTIONS

This study primarily evaluates computational efficiency and accuracy. Power consumption analysis, robustness to varying lighting conditions, and real-world deployment in physical smart trash bins were beyond the current scope but remain essential future research directions. Therefore, in the future, the models will be tested on real-world datasets under conditions such as low light, image noise, and environmental factors. Subsequently, the classification system will be integrated into smart trash bins, and fully integrated embedded systems will be designed and tested, including sensors, processors, user interfaces, and data transmission systems for remote monitoring and control. Classification models will be coordinated with IoT systems to collect larger datasets, analyze user behavior, and optimize waste collection and sorting processes. Developing unsupervised or semi-supervised learning will reduce dependence on costly labeled data. Small-scale pilot systems will be deployed in real urban areas to evaluate practical performance, stability, and adaptability of the model. User feedback will be collected to refine the waste management and treatment system in smart cities.

VI. CONCLUSION

The machine learning models proposed in this study, namely ConvNeXt and PrQ-CNN, have demonstrated high accuracy in classifying Non-recyclable and Recyclable solid waste, achieving 98% and 92%, respectively. This study also proposed techniques such as pruning and quantization to reduce model size for integration onto embedded devices with limited hardware resources, such as Raspberry Pi. Experimental results show that both models have achieved not only superior classification efficiency but are also suitable for integration into smart trash bin systems. This study highlights the potential for application in waste management and treatment classification systems, contributing to improved efficiency, reduced errors, and cost optimization. Compared to conventional models such as VGG16, ResNet, or MobileNet, the proposed models in this study offer a good balance between accuracy, size, and inference time, meeting the requirements of embedded systems. It also demonstrates the potential for applying artificial intelligence technology in waste classification, which not only enhances processing efficiency but also promotes the development of smart, sustainable waste management systems, contributing to environmental protection initiatives against climate change and improving quality of life in urban areas.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Pham Trung Kien.
Methodology: Pham Trung Kien.
Investigation: Pham Trung Kien.
Discussion of results: Pham Trung Kien.
Writing – Original Draft: R. Pham Trung Kien.
Writing – Review and Editing: Pham Trung Kien.
Resources: Pham Trung Kien.
Supervision: Pham Trung Kien.
Approval of the final text: Pham Trung Kien.

VIII. DATA AVAILABILITY STATEMENT

Data supporting reported results can be downloaded at: *https://www.kaggle.com/techsash/waste-classification-data* (accessed on 09 January 2026).

IX. REFERENCES

[1] T. Kumari and A. S. Raghubanshi, "Waste management practices in the developing nations: challenges and opportunities," in Waste Management and Resource Recycling in the Developing World, Elsevier, 2023, ch. 33, pp. 773–797, doi: 10.1016/B978-0-323-90463-6.00017-8.

[2] A. Lakhouit, "Revolutionizing urban solid waste management with AI and IoT: A review of smart solutions for waste collection, sorting, and recycling," Results in Engineering, vol. 25, 2025, doi: 10.1016/j.rineng.2025.104018.

[3] A. Addas, M. N. Khan, and F. Naseer, "Waste management 2.0: Leveraging Internet of Things for an efficient and eco-friendly smart city solution," PLoS One, vol. 19, 2024, doi: 10.1371/journal.pone.0307608.

[4] F. Fotovvatikhah, I. Ahmedy, R. M. Noor, and M. U. Munir, "A systematic review of AI-based techniques for automated waste classification," Sensors, vol. 25, 2025, doi: 10.3390/s25103181.

[5] Z. Zhang, C. Zhonghao, J. Zhang, Y. Liu, L. Chen, M. Yang, A. I. Osman, M. Farghali, E. Liu, D. Hassan, I. Ihara, K. Lu, D. W. Rooney, and P. S. Yap, "Municipal solid waste management challenges in developing regions: A comprehensive review and future perspectives for Asia and Africa," Science of the Total Environment, vol. 930, 2024, doi: 10.1016/j.scitotenv.2024.172794.

[6] J. Ma, Z. Chen, J. Li, L. Liu, Q. Yang, Y. Wan, X. Lu, Y. Zhang, and X. Fei, “Concurrent aerobic methane oxidation and biodegradation of waste in shallow layer of landfill during aeration,” Journal of Environmental Management, vol. 381, 2025, doi: 10.1016/j.jenvman.2025.125282.

[7] M. M. Abo-Zahhad and M. Abo-Zahhad, “Real-time intelligent garbage monitoring and efficient collection using YOLOv8 and YOLOv5 deep learning models for environmental sustainability,” Scientific Reports, vol. 15, 2025, doi: 10.1038/s41598-025-99885-x.

[8] I. Ranjbar, Y. Ventikos, and M. Arashpour, “Deep learning-based construction and demolition plastic waste classification by resin type using RGB images,” Resources, Conservation and Recycling, vol. 212, 2025, doi: 10.1016/j.resconrec.2024.107937.

[9] M. C. Bello, D. B. E. Padilla, C. M. Morales, W. C. Francisco, C. V. M. Vega, C. M. Duarte, Y. E. Alcocer, and D. E. G. Valencia, “Convolutional neural network models in municipal solid waste classification: Towards sustainable management,” Sustainability, vol. 17, 2025, doi: 10.3390/su17083523.

[10] P. Wang, J. Xiao, R. Liu, X. Qiang, X. Duan, and C. Liang, “A hybrid classification and evaluation method based on deep learning for decoration and renovation waste in view of recycling,” Waste Management, vol. 191, pp. 1–12, 2025, doi: 10.1016/j.wasman.2024.10.027.

[11] J. Son and Y. Ahn, “AI-based plastic waste sorting method utilizing object detection models for enhanced classification,” Waste Management, vol. 193, pp. 273–282, 2025, doi: 10.1016/j.wasman.2024.12.014.

[12] S. Tiwari, S. Bisht, and K. Sharma, “Intelligent waste management using WasteIQNet with hierarchical learning and meta-optimization,” IEEE Access, 2025, doi: 10.1109/ACCESS.2025.3574095.

[13] J. Tanvir, M. S. K. Tanzir, B. K. Paul, M. Morshed, “TrashNeXt: Classification of recyclable water pollutants using deep transfer learning method,” Case Studies in Chemical and Environmental Engineering, vol. 11, 2025, doi: 10.1016/j.cscee.2024.101073.

[14] G. Mazanov, A. Iliushina, S. Nesteruk, A. Pimenov, A. Stepanov, and N. Mikhaylova, “Hyperspectral data-driven solid waste classification,” IEEE Access, vol. 13, pp. 46659–46672, 2025, doi: 10.1109/ACCESS.2025.3551097.

[15] M. S. B. Islam, M. S. I. Sumon, M. E. Majid, S. B. A. Kashem, M. Nashbat, A. Ashraf, A. Khandakar, A. K. A. Kunju, M. Hasan-Zia, and M. E. H. Chowdhury, “ECCDN-Net: A deep learning-based technique for efficient organic and recyclable waste classification,” Waste Management, vol. 193, pp. 363–375, 2025, doi: 10.1016/j.wasman.2024.12.023.

[16] H. T. N. Le and H. Q. T. Ngo, “Application of the vision-based deep learning technique for waste classification using the robotic manipulation system,” International Journal of Cognitive Computing in Engineering, vol. 6, pp. 391–400, 2025, doi: 10.1016/j.ijcce.2025.02.005.

[17] A. M. Vukicevic, M. Petrovic, N. Jurisevic, M. Djapan, N. Knezevic, A. Novakovic, and K. Jovanovic, “Versatile waste sorting in small batch and flexible manufacturing industries using deep learning techniques,” Scientific Reports, vol. 15, 2025, doi: 10.1038/s41598-025-87226-x.

[18] M. M. Taye, “Theoretical understanding of convolutional neural networks: Concepts, architectures, applications, and future directions,” Computation, vol. 11, 2023, doi: 10.3390/computation11030052.

[19] J.-H. Luo, H. Zhang, H.-Y. Zhou, C.-W. Xie, J. Wu, and W. Lin, “ThiNet: Pruning CNN filters for a thinner net,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, 2019, doi: 10.1109/TPAMI.2018.2858232.

[20] S. I. Young, Z. Wang, D. Taubman, and B. Girod, “Transform quantization for CNN compression,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 9, pp. 5700–5714, 2022, doi: 10.1109/TPAMI.2021.3084839.

[21] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A ConvNet for the 2020s,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 11966–11976, doi: 10.1109/CVPR52688.2022.01167.

[22] C. Shi, C. Tan, T. Wang, and L. Wang, “A waste classification method based on a multilayer hybrid convolution neural network,” Applied Sciences, 2021, doi: 10.3390/app11188572.

[23] M. Malik, S. Sharma, M. Uddin, C. L. Chen, C. M. Wu, P. Soni, and S. Chaudhary, “Waste classification for sustainable development using image recognition with deep learning neural network models,” Sustainability, 2022, doi: 10.3390/su14127222.

[24] N. Nnamoko, J. Barrowclough, J. Procter, “Solid Waste Image Classification Using Deep Convolutional Neural Network,” Infrastructures, 2022, doi: 10.3390/infrastructures7040047

[25] Th. Vadivel, Kumar Barathi, G. Arulkumaran, M. B. Anand and C. Cherubini, “Wastewater Management Using a Neural Network-Assisted Novel Paradigm for Waste Prediction from Vermicomposting,” in Water, MDPI, 2024, vol. 33, issue. 23, doi: 10.3390/w16233450

[26] R. Alam, I. Darmawan, P. A. Alda, and F. M. N. Sugiartana, “WasteInNet: Deep learning model for real-time identification of various types of waste,” Cleaner Waste Systems, vol. 10, 2025, doi: 10.1016/j.clwas.2024.100198.