A Novel Approach for Waste Classification in Edge Devices

Tuteja, Anvi

Science Suncity School Gurgaon, India
0009-0003-8791-5464

DOI: 10.29322/IJSRP.14.10.2024.p154XX

Paper Received Date: 13th September 2024
Paper Acceptance Date: 14th October 2024
Paper Publication Date: 22nd October 2024

***Abstract*—**Waste management is a growing global concern, with over 2 billion tons of waste produced annually, of which nearly 10% is hazardous. Traditional methods of waste sorting are inefficient, leading to significant environmental and health impacts. Recent advancements in computer vision and deep learning have opened opportunities for automated waste classifi- cation, which can significantly enhance the efficiency of recyclingprocesses. This paper investigates the use of convolutional neuralnetworks (CNNs) for automated waste classification. The main focus of the research is to identify CNN models that have higher classification accuracy and can be deployed on edge devices with limited computational capacity. The study comprise evaluation ofthe models across various waste management datasets, ranking them based on their F1 scores for classification accuracy. Addi- tional considerations were applied for computational efficiency, measured in GFLOPS, and the number of model parameters to identify CNN models that require fewer resources. The aimis to identify models that balance high accuracy with low computational demands, enabling real-time waste classification. This study advances waste management practices by offeringefficient, scalable solutions that can be implemented on edge devices for sustainable waste sorting.

***Index Terms*—**convolutional neural networks, waste classifica- tion, lightweight, deep learning

I. Introduction

Waste is becoming a significant global problem. Every year, we dump over 2 billion tons of waste on our planet[1], of which nearly 10% is hazardous waste [5]. The waste we generate causes all types of pollution including soil, water, and air leading to untimely deaths, diseases, and permanent impacts on our environment [17]. According to The World Count Report, waste production is increasing substantially and if this problem isn't solved, we may need an extra planet of the size of Earth to provide resources and absorb our waste.

In recent times, the waste has grown not only in volume but also in its impact on the environment. According to Statista[2] e-waste accounted for 62 million tons of waste during 2022 and it is expected to grow to 80 million tons by 2030. The World Health Organization's report on e-waste[3] highlights that less than 20% of global e-waste is formally documented and collected leading to a significant impact on our planet. E-wastegenerally contains hazardous materials such as lead, cadmium,and mercury, which cause adverse impacts on our soil and alsoto public health.

While the best approach to reduce waste is by minimizing itsgeneration, it's not easy due to an increase in population, rapidurbanization, and growing consumerism. Recycling has been one of the most effective approaches for waste management, but in many developing countries, informal recycling practices also expose vulnerable populations especially young children to toxic substances. Further, the advanced developments in product design and usage have led to enhanced circularitywherein the use of our resources can be extended and op- timized. There is a growing need to develop safe, efficient, and effective approaches to the classification of waste in various categories for proper disposal, recycling, and enablingcircularity.

Traditional recycling facilities sort waste material through a series of filters largely based on size. However, the best approach to sorting waste is closer to its generation. A few European countries have developed regulations for mandatory waste separation, but it is still not a general practice globally. Several municipal corporations and local governments have developed guidelines and rules for the separation and recycling of waste.

With rapid advancements in Computer Vision technologies, it is possible to improve waste classification across generation, collection, and disposal points. Researchers around the world have used various machine learning and deep learning tech- niques to develop systems that can detect and classify waste material with reasonable accuracy [4], [6], [18]. This research study aims to evaluate the existing approaches and develop new approaches for enhancing accuracy and improving effi- ciency while reducing the need for computational resources.

One of the limiting factors for deploying AI solutions in the past has been the cost and availability of computational capacity closer to the endpoint. However, recent developments in affordable and efficient microprocessors like Raspberry Pi have given rise to building edge devices at low cost, which can

[1]https://www.theworldcounts.com/challenges/planet-earth/waste/global-waste-problem

[2]https://www.statista.com/topics/3409/electronic-waste-worldwide/#editorsPicks

[3]https://www.who.int/news-room/fact-sheets/detail/electronic-waste-28e-waste29

be deployed in large numbers. These devices can be deployed closer to the waste management points like bins and have the potential to substantially reduce the cost of waste sorting and management.

The objective of this research article is to determine the relative ranking of various convolutional neural networks for the classification of waste materials. The models will be ranked based on their F1 score, depicting the overall performance. Since computer vision-based waste classification can be applied in a variety of use cases ranging from waste classification at source (domestic and industrial) to classification at municipal waste management areas, different models may be suitable for different applications. In addition, the focus will be on finding image classification models that give better F1 scores with less number of computations (GFLOPS) and model parameters so that they can be deployed in edge devices. The following are the objectives of this study:

- Rank various deep-learning models for waste material classification based on their F1 scores, considering different application contexts like domestic, industrial, and municipal waste management.
- Find the best-performing image classification model that requires less number of computations (GFLOPS) and model parameters.

II. Literature Review

Given the growing nature of the problem, several studies have been done to understand and improve waste classification strategies. This section attempts to summarize the information available so far and the work done by various researchers.

A. Waste Classification Challenges

Waste classification is one of the most impactful components of effective waste management leading to a higher degree of recycling and enabling circularity. Traditional methods of waste separation are largely manual leading to error and safety hazards. Past studies indicate that manual sorting methods can only achieve 50-70% accuracy and are highly inefficient [13]. This will lead to contamination and increased operational costs. As the volume of waste continues to grow and its composition becomes more complex, there is a need to develop automated systems to classify and separate waste material at reduced cost and higher efficiency.

B. Use of Artificial Intelligence in Waste Classification

Several researchers have proposed the use of AI systems for the classification of waste using neural networks and computer vision [1]. Initially, traditional Machine Learning (ML) models like Support Vector Machine (SVM) and Decision Tree methods have been suggested to classify waste material[4]. However, developments in Deep Learning (DL) models like Convolutional Neural Networks (CNN) enabled high-performance processing of images for recognition and classification. CNNs typically have a high number of layers and can learn features and patterns from input images leading to faster training and increased recognition accuracy.

There have been several studies that have used a range of pre-trained CNN models to classify waste material. Interestingly, these studies have indicated varied accuracy levels.

[4]https://github.com/Morris88826/waste item classification

Many researchers have attempted varied approaches for increasing the accuracy and F1 scores of these classification models [3]. Over time, the research has largely moved to using deep transfer learning models for improving accuracy and efficiency. Many pre-trained deep learning models using ImageNet [9] have resulted in relatively high classification performance. However, studies also indicate that the models suffer from overfitting and category imbalance challenges. The results also suggest that a combination of supervised learning with CNN can deliver exceptional performance.

One of the studies indicates that models like VGG16 [14] and VGG19 [14] with few epochs have been found to deliver accuracy levels above 90% [8]. Another study advocates the use of XceptionNet [7] as a good choice for delivering better accuracy levels. According to a study, larger models like ResNet50 [10] and ResNet152 [10] will offer higher accuracy levels.

A literature review study by Abood, Isara Nasir [12] summarizes the accuracy levels of various machine learning and deep learning models. While there are several comparative studies available, these studies compare the performance of various models using different datasets, trained at different epoch levels and with or without using data augmentation techniques. Further, it is important to compare the F1 scores with consistent datasets to arrive at a reasonable conclusion.

C. Hybrid Approaches

Transfer learning has been gaining popularity in waste classification research. Researchers have advocated the use of a pre-trained model with fine-tuning using domain-specific datasets. This approach offers higher efficiency and faster learning times. Further, another research suggests that the use of CNN along with the SVM learning technique can provide better results. According to this approach, CNN has been used for feature extraction and SVM has been used for classificationoffering enhanced performance.

Kumsetty, et al.'s novel approach of using Quantum TransferLearning offers promising results [11]. This research suggests the use of a quantum circuit network in conjunction with classical transfer learning to enhance the model performance. The findings indicate a 10.84% improvement in performance and a 27.4% decline in training time.

D. Data Augmentation

The accuracy of a computer vision model in real life will greatly depend on the variety of data it has been trained on. There are a limited number of datasets available, which canbe used to train a deep learning model. TrashNet [16] is one of the most widely used sources by the researchers. One of the researchers has also undertaken a crowdsourcing project to collect labeled data of different types of trash. However, the size of such datasets remains relatively small.

Data augmentation is a popular technique to increase the size of the dataset, which can be used for training the deep learning models. One of the basic techniques is to apply various transformations like random rotation, flip, and zoom to

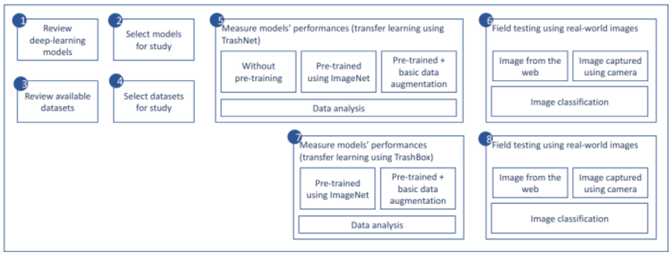

Fig. 1: Overall Methodology

existing images. While the resultant dataset can be very large in terms of the number of images, it lacks the diversity and uniqueness of features.

The recent advances in Generative Adversarial Networks (GAN) offer new opportunities for generating synthetic images to augment the existing datasets. According to a study by Alsabei, et al. [2], the use of StyleGAN indicated different sensitivity levels to the accuracy for a range of augmented images. The research covered studying the impact of generated images on the performance of four deep learning models (VGG16, Xception, Inception V3 [15], ResNet50). The findings indicate that all the models performed better without adding generated images. While the models based on VGG16 and Xception remained insensitive to the addition of generated images, the ResNet50 and Inception V3-based models showed deterioration in their performance with the addition of generated images.

III. Methodology

This study involved the following phases to achieve the two specified objectives as illustrated in Figure 1.

A. Review available deep-learning models

A large spectrum of available deep-learning models were reviewed. This stage included understanding the model architecture, number of layers, size of the model, computational complexity, top accuracy and transfer learning capabilities.

Number of layers in a model indicates its depth and is typically an indicator of the model's ability to learn complex patterns. A model with a larger number of layers can learn more complex patterns but it comes with the cost of learning time and computational resources.

The size of the model indicates the number of parameters it contains. The larger models have a better capacity to learn and can differentiate the intricacies between the learning datasets. However, it adds to the cost of computational resources. Computational complexity indicates how many billion floating-point operations can a model typically perform in one second. It is measured in Giga Floating Point Operations/second (GFLOPs). A model with higher GFLOPs indicates that it can efficiently perform complex computations.

Top Accuracy or Top-1 Accuracy (also indicated as acc@1) is a measure that indicates the percentage of instances wherein the model's best-predicted category is the same as the right category. This is a measure that the model's first (or most confident) prediction is correct. The higher the Acc@1 score of a model, the more accurate it is likely to be.

TABLE I: CNN models used for waste classification task with their ImageNet accuracy, number of parameters and computations required in GFLOPS.

Model	ImageNet Acc@1	Params	GFLOPS	Model Category
alexnet	56.52	61.1M	0.71	Small
efficeintnet b0	77.69	5.3M	0.39	Small
efficeintnet b7	84.12	66.3M	37.75	Large
mobilenet v2	71.88	3.5M	0.30	Small
mobilenet v3 large	74.04	5.5M	0.22	Small
mobilenet v3 small	67.67	2.5M	0.06	Small
resnet18	69.76	11.7M	1.81	Medium
resnet34	73.31	21.8M	3.66	Medium
resnet50	76.13	25.6M	4.09	Medium
resnet101	77.37	44.5M	7.80	Medium
resnet152	78.31	60.2M	11.51	Large
shufflenet v2 x0 5	60.55	1.4M	0.04	Small
shufflenet v2 x2 0	76.23	7.4M	0.58	Small
vgg16	71.59	138.4M	15.47	Large
vgg13	69.93	133.0M	11.31	Large
vgg19	74.22	143.7M	19.63	Large

The deep learning models are generally pre-trained on a larger number of generic datasets. The transfer learning capability suggests that a model can augment its learning of pre-trained data from additional domain-specific data. This ability is key to making the model efficient in recognizing patterns for a specific type of dataset.

Initially, a total of 115 deep-learning models were evaluated for their relevance to this study. Based on a high-level review of available data, a short list of 16 models was created for further analysis. Table I compares the selected attributes of various deep learning models evaluated for this study.

B. Model selection

Based on the key attributes discussed and the comparative performance of these models in previous studies, 16 deep learning models (see Table I) have been selected for further analysis. These models are categorized into three groups according to their computational cost, measured in GFLOPS (Giga Floating Point Operations per Second), which indicates the number of floating-point operations required to process a single image. GFLOPS provides a useful metric for comparing models based on their computational complexity, offering insight into their suitability for deployment on various devices, particularly those with limited computational resources. Models with 0 to 1 GFLOPS are classified as small, those with 1 to 10 GFLOPS as medium, and those requiring more than 10 GFLOPS fall into the large category.

C. Review and selection of datasets

Most of the deep learning models selected for this study have been pre-trained on ImageNet, which is the most popular and widely used dataset for image recognition. The ImageNet [9] contains over 14 million labeled images across thousands of categories. While the ImageNet pre-trained models are highly effective start points, their performance can be

(a) Plastic (b) Paper (c) Metal (d) Medical Waste (e) Glass (f) Cardboard (g) E-Waste

Fig. 2: Sample images of waste from different categories present in TrashBox dataset.

significantly enhanced by fine-tuning through domain-specific (waste-related) image datasets. For the purpose of this study, several datasets for fine-tuning the models were evaluated. These datasets have been summarized in Table II.

TrashNet, while limited in size and having category imbalance has been used by various researchers. It offers good training images with adequate lighting but may not represent the data as presented in real-life use cases. The Waste Classification dataset is large in size but contains very few classes. TACO contains images closer to what may be found in real-life use cases but has a much smaller number of images. TrashBox and Garbage Classification are both larger datasets and are suitable for this study. Finally, TrashNet and TrashBox datasets were selected for this study because of their relevance to the potential use cases and number of categories. It is worth noting that TrashBox contains several sub-categories and there are a few inconsistencies in categorization. A few examples of TrashBox are shown in Figure 2. For example, cigarette butt is a sub-category of Plastic category. Similarly, TrashNet contains a category named – Trash, which contains random trash images and can be misleading for model training. However, such inconsistencies are not significant and will not likely have a major impact on the outcome of the current study. The dataset statistics of TrashBox and TrashNet are shown in Figure 3.

D. Measuring Models' Performance

To measure the performance of the selected models, each of the models was evaluated for a number of performance metrics. Several past studies have limited their analysis to measuring the accuracy of the model. However, this can provide misleading results due to category imbalance. The current study evaluated each model on several performance metrics including Accuracy, Precision, Recall, and F1 Score.

Accuracy measures the overall proportion of correct predictions made by a model across all classes. Accuracy provides a general sense of a model's correctness.

$$Accuracy = \frac{TruePositives + TrueNegatives}{TotalPredictions} \quad (1)$$

It is a simple metric for balanced datasets but can be misleading in the case of imbalanced categories. Further, it does not provide insights into the model's performance in individual classes. Precision measures the proportion of true positive predictions among all positive predictions made by the model. It's a good measure to understand the model's ability to avoid false positives.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (2)$$

Precision ignores false negatives and is an important metric to consider when the cost of false positives is high.

Recall measures the proportion of actual positive instances that are correctly identified by the model. It indicates the model's ability to find all the relevant cases.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (3)$$

Recall ignores false positives and is an important metric to consider when the cost of false negatives is high.

F1 Score the harmonic mean of precision and recall, providing a balanced metric that considers both false positives and false negatives.

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

F1 Score is a useful metric when there are category imbalances in the dataset and accuracy alone can not be considered a reliable metric.

IV. Experiment setup

All the selected CNN models are trained with TrashNet and TrashBox datasets. The images of these datasets are first preprocessed to 224x224 pixels. As the TrashNet dataset does not have any specific train-test split, a custom train-test and validation split was created by randomly assigning images. The train set split of TrashNet dataset contains 2,020 images and 252 for test and val each. For the TrashBox the train split contains 13,377, val split contains 2,580 and test split contains 1,828 images. All the CNN models are trained for 100 epochs with batch size 32. The models are optimized with Adam optimizer with a learning rate of 0.001.

V. Results and Discussion

A. Model Evaluation on TrashNet

Initially, the CNN models are first trained with the TrashNet dataset as the dataset contains less number of images. First, the models are trained solely on TrashNet without any pertaining. Next, all CNN models are initialized with ImageNet trained weights and later fine-tuned. Additionally, experiments are also conducted with and without inclusion of data augmentation.

TABLE II: Dataset on Waste classification

Dataset Name and description	Categories	Number of images
TrashNet: The dataset was created by taking pictures of the objects on a white posterboard in sunlight/ room light.	6 (cardboard, glass, metal, paper, plastic, and trash)	2,527
Waste Classification Data: This dataset contains 22,500 images of organic and recyclable objects.	2 (organic and recyclable)	22,500
TACO (Trash Annotations in Context): An open image dataset of waste in the wild. It contains photos of litter taken in diverse environments. The annotations are done by the community collaborators.	60 categories (28 super categories)	1,500
TrashBox: A large dataset of waste images including modern waste categories created by labelling Internet scraped images.	7 (glass, plastic, metal, e-waste, cardboard, paper, medical waste)	17,785
Garbage Classification: Augmented dataset using existing data sources and web-scraped images.	12 (paper, cardboard, biological, metal, plastic, green-glass, brown-glass, white-glass, clothes, shoes, batteries, trash)	15,150
WasteNet: Limited information is available in the public domain.		3 million

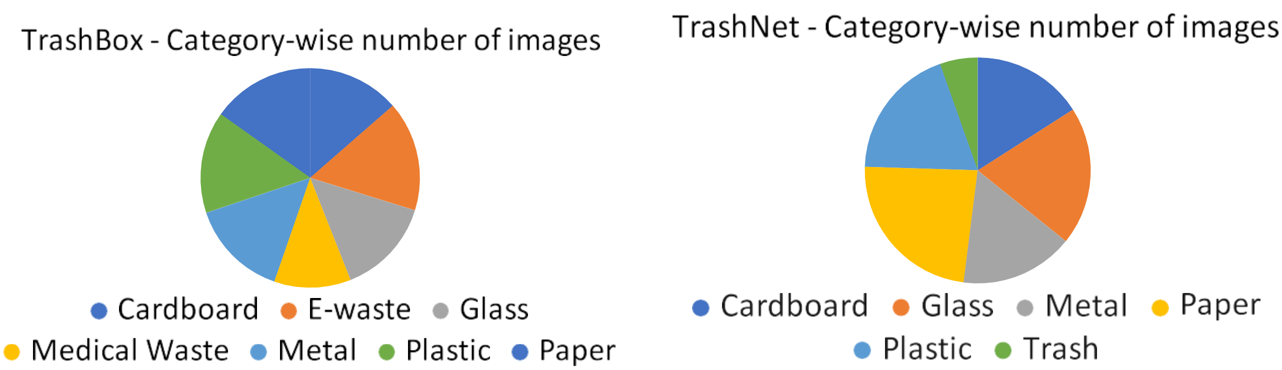

Fig. 3: Waste category wise image distribution of TrashNet and TrashBox dataset.

The CNN models were evaluated for Accuracy, Precision, Recall, and F1 Score using the TrashNet dataset and the results are shown in Table III. The models were evaluated under the following different scenarios:

- Without Pre-Training – in this scenario, the models were evaluated with no pre-training or pre-defined weights.
- Pre-Trained Using ImageNet – in this scenario, the models were evaluated after being trained on the ImageNet dataset.
- Pre-Trained with augmentation – in this scenario, the models were evaluated after being trained on ImageNet and further augmenting the dataset with basic augmentation techniques like rotation, flip and zoom.

Following observations are made from results presented in Table III:

- **Small Models (0 to 1 GFLOPS)**: AlexNet, EfficientNet_b0, Mobilenet_v2, Mobilenet_v3 (large and small), and ShuffleNet variants fall under the small models category, with GFLOPS ranging from 0.04 to 0.71. Without pre-training, models in this category generally show lower performance. For instance, mobilenet_v3_small has a 69.67% F1 score, while mobilenet_v2 achieves 80.57%. This indicates that these models may struggle to generalize well without pre-training on large datasets. However, pre-training with ImageNet leads to a significant improvement in performance. AlexNet, for example, jumps to an F1 score of 91.45%, and mobilenet_v2 reaches 94.24%, and EfficientNet_b0 performs excep-

TABLE III: Performance of different CNN models trained and tested using TrashNet dataset.

Model	GFLOPS	Without Pre Training				Pre Trained with ImageNet				Pre Trained with ImageNet + Augmentation			
		Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
alexnet	0.71	75.10	75.55	75.10	75.06	91.50	91.66	91.50	91.45	91.90	91.89	91.90	91.84
efficeintnet b0	0.39	77.47	78.07	77.47	77.19	95.45	95.53	95.45	95.47	96.05	96.10	96.05	96.05
efficeintnet b7	37.75	80.24	80.96	80.24	80.24	93.08	93.07	93.08	93.05	93.28	93.38	93.28	93.28
mobilenet v2	0.30	80.83	80.76	80.83	80.57	94.27	94.37	94.27	94.23	95.65	95.71	95.65	95.65
mobilenet v3 large	0.22	75.49	75.96	75.49	75.06	94.66	94.69	94.66	94.62	95.65	95.80	95.65	95.67
mobilenet v3 small	0.06	69.96	70.88	69.96	69.67	93.28	93.54	93.28	93.27	94.27	94.30	94.27	94.24
resnet18	1.81	82.21	82.91	82.21	82.00	92.49	92.70	92.49	92.45	93.68	93.80	93.68	93.69
resnet34	3.66	78.85	79.27	78.85	78.65	94.27	94.27	94.27	94.23	94.66	94.75	94.66	94.67
resnet50	4.09	78.66	78.72	78.66	78.23	94.86	94.96	94.86	94.83	95.85	95.81	95.85	95.78
resnet101	7.80	76.68	76.95	76.68	76.24	94.86	94.88	94.86	94.84	96.05	96.10	96.05	96.02
resnet152	11.51	78.66	80.02	78.66	78.03	96.05	96.05	96.05	96.04	96.64	96.68	96.64	96.63
shufflenet v2 x0 5	0.04	73.91	73.99	73.91	73.66	84.58	85.98	84.58	82.39	83.20	85.15	83.20	80.40
shufflenet v2 x2 0	0.58	76.48	77.35	76.48	76.21	87.94	87.56	87.94	85.69	87.75	88.86	87.75	84.73
vgg16	15.47	72.92	72.82	72.92	72.50	93.48	93.68	93.48	93.46	94.47	94.63	94.47	94.46
vgg13	11.31	76.48	76.64	76.48	76.39	93.28	93.35	93.28	93.19	94.66	94.69	94.66	94.66
vgg19	19.63	76.28	77.25	76.28	75.97	91.30	91.42	91.30	91.32	94.07	94.25	94.07	94.10

tionally well at 95.05%. These models, despite their low computational cost, perform competitively when pre-trained. Additionally, pre-training with augmentation further boosts the performance of these models. For example, mobilenet_v2 reaches an F1 score of 95.65%, and mobilenet_v3_small performs at 94.24%. This demonstrates that even small GFLOPS models can achieve strong results when combined with proper pre-training and data augmentation techniques. In the small category models mobilenet_v3_large archives the highest F1 score of 95.67%.

- **Medium Models (1 to 10 GFLOPS)**: ResNet18, ResNet34, ResNet50, and ResNet101 fall within the medium models category, with GFLOPS ranging from 1.81 to 7.80, striking a balance between computational complexity and performance. Even without pre-training, these models outperform smaller ones, with ResNet18 achieving the highest accuracy 82.21% and F1 score 82.00%, followed by ResNet18 with a 78.65% F1 score. Pre-training with ImageNet significantly boosts their performance, with ResNet50 reaching an F1 score of 94.83% and ResNet101 achieving 94.84%, demonstrating their efficiency despite moderate GFLOPS. Further improvement is seen with pre-training and augmentation, where ResNet50 achieves a 95.78% F1 score and ResNet101 reaches 96.02%, highlighting that medium GFLOPS models provide an optimal balance between computational efficiency and high performance when enhanced with pre-training and augmentation.
- **Large Models (Above 10 GFLOPS)**: ResNet152, VGG13, VGG16, VGG19, and EfficientNet_b7 are classified as large models, with GFLOPS ranging from 11.31 to 37.75. Even without pre-training, these models perform relatively well compared to smaller models, with EfficientNet_b7 achieving an F1 score of 80.96% and ResNet152 reaching 80.02%, though their computational cost is significantly higher. Pre-training with ImageNet significantly boosts their performance, with ResNet152 achieving an F1 score of 96.04% and EfficientNet_b7 reaching 93.05%, demonstrating the substantial benefits of pre-training that justify their high computational cost. Further improvements are seen with pre-training and augmentation, as ResNet152 attains the highest F1 score 96.63% and EfficientNet_b7 reaches 93.28%, highlighting that large models, despite being computationally expensive, can deliver superior performance when effectively pre-trained and fine-tuned with augmentation techniques.

Small models excel in computational efficiency and performance, making them ideal for devices with limited resources, especially when pre-trained. Medium models strike a better balance, with ResNet50 and ResNet101 emerging as top performers due to their moderate computational costs and enhancements from pre-training and augmentation. Large models, such as ResNet152 and EfficientNet_b7 variants, achieve the highest performance but require significant GFLOPS, making them suitable for scenarios with ample computational resources.

B. *Model Evaluation on TrashBox*

TABLE IV: The performance of the models with less computational cost was trained and tested with the TrashBox dataset.

Model	GFLOPS	TrashBox			
		Accuracy	Precision	Recall	F1
efficeintnet b0	0.39	92.43	92.47	92.43	92.43
mobilenet v2	0.30	91.29	91.39	91.29	91.28
mobilenet v3 large	0.22	91.47	91.49	91.47	91.47
mobilenet v3 small	0.06	88.16	88.17	88.16	88.14

The TrashNet dataset, consisting of only 2,527 images, poses a risk of models achieving higher F1 scores during testing but potentially underperforming in real-world scenarios. To assess the impact of dataset size and diversity, the small category models were evaluated using the larger TrashBox dataset, which contains 17,785 images. The performance of these models are summarized in Table IV. The objective of this experiment is to identify the best model with minimal computational costs. However, the ShuffleNet variants have been excluded from this analysis due to their low performance.

The performance evaluation of models on the TrashBox dataset highlights the effectiveness of lightweight architectures in achieving competitive accuracy while maintaining low computational costs. EfficientNet_b0 stands out as the top performer, achieving an accuracy of 92.43% with closely aligned precision and recall metrics, indicating its robustness and reliability. Both MobileNet_v2 and MobileNet_v3_large also demonstrate solid performance with accuracies of 91.29% and 91.47%, respectively, effectively balancing precision and recall. The GFLOPS values reveal a compelling trade-off between performance and computational demands, with MobileNet_v3_small showing lower accuracy at 88.16% despite its minimal computational cost of 0.06 GFLOPS, suggesting a potential trade-off when prioritizing efficiency over performance. These results imply that EfficientNet_b0 is the most suitable for applications requiring high accuracy and reasonable resource usage, while the MobileNet variants are viable for edge devices where computational resources are limited. Overall, the analysis underscores the importance of selecting models that deliver reliable performance metrics in real-world scenarios, ensuring that deep learning applications remain feasible and effective in resource-constrained environments.

VI. Conclusion

This paper summarizes the results of a study that was performed to evaluate the suitability of various deep-learning models in detecting and classifying waste material. While most models tend to offer good performance with pre-training, the model performance does not improve significantly through basic data augmentation techniques. One of the objectives of this study was to categorize models in terms of their suitability for various use cases. Since the smaller models like EfficientNet B-0, MobileNet V2, and MobileNet V3 Large perform reasonably well despite their small footprint and limited computational resource requirements, these models should be preferred for deployment in domestic use cases like smart bins or robotic arm separators. These models can work efficiently on smaller microprocessors like Raspberry Pi or equivalent. On the other hand, larger models like ResNet152 and ResNet101 offer better performance but can be resource-intensive. Therefore, these models should be deployed in industrial use cases like municipal waste management facilities. The second objective of the study was to evaluate the impact of data augmentation on model performance. The results strongly suggest that the basic data augmentation techniques like rotation, flip, and zoom do not lead to significant improvement in the model performance. However, the size of the dataset, its composition, and category balance play a much larger role. The results highlight an important insight. While the theoretical performance (using validaton data by splitting the training datasets in train/test) of models trained using a smaller dataset TrashNet appears higher than those trained using TrashBox, the real-world performance of TrashBox was found to be much better. This highlights the importance of using larger, more balanced, and real-life representative datasets for training models to be deployed for waste classification.

References

[1] Olugboja Adedeji and Zenghui Wang. Intelligent waste classification system using deep learning convolutional neural network. *Procedia Manufacturing*, 2019.

[2] Amani Alsabei, Ashwaq Alsayed, Manar Alzahrani, and Sarah Al-Shareef. Waste classification by fine-tuning pre-trained cnn and gan. *International Journal of Computer Science & Network Security*, 21(8):65–70, 2021.

[3] Zian Md Afique Amin, Khan Nasik Sami, and Raini Hassan. An approach of classifying waste using transfer learning method. *International Journal on Perceptive and Cognitive Computing*, 7(1):41–52, 2021.

[4] Sam Azadi and Ayoub Karimi-Jashni. Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of fars province, iran. *Waste management*, 48:14–23, 2016.

[5] KO Babaremu, SA Okoya, E Hughes, B Tijani, D Teidi, A Akpan, J Igwe, S Karera, M Oyinlola, and ET Akinlabi. Sustainable plastic waste management in a circular economy. *Heliyon*, 8(7), 2022.

[6] Kamakhya Bansal and Ashish Kumar Tripathi. Wastenet: A novelmulti-scale attention-based u-net architecture for waste detection in uav images. *Remote Sensing Applications: Society and Environment*, 2024.

[7] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[8] Ying-Chien Chu, Chen Huang, Xiaodan Xie, Bohai Tan, Shyam Kamal, and Xiaogang Xiong. Multilayer hybrid deep-learning method for waste classification and recycling. *Computational Intelligence and Neuroscience*, 2018, 2018.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[11] Nikhil Venkat Kumsetty, Amith Bhat Nekkare, Sowmya Kamath S., and Anand Kumar M. Trashbox: Trash detection and classification using quantum transfer learning. *2022 31st Conference of Open Innovations Association (FRUCT)*, pages 1–6, 2022.

[12] Israa Nasir and Ghaidaa Abdul Aziz Al-Talib. Waste classification using artificial intelligence techniques: literature review. 2023.

[13] Khan Nasik Sami, Zian Md Afique Amin, and Raini Hassan. Waste management using machine learning and deep learning algorithms. *International Journal on Perceptive and Cognitive Computing*, 2020.

[14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[15] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[16] Gary Thung and Mingxiang Yang. Classification of trash for recyclability status. 2016.

[17] Stephanie L Wright and Frank J. Kelly. Plastic and human health: A micro issue? *Environmental science & technology*, 51 12:6634–6647, 2017.

[18] Wanjun Xia, Yanping Jiang, Xiaohong Chen, and Rui Zhao. Application of machine learning algorithms in municipal solid waste management: A mini review. *Waste Management & Research*, 40(6):609–624, 2022.