Performance Comparison of Pretrained Deep Learning Models for Landfill Waste Classification

Hussein Younis[1], Mahmoud Obaid[2]
Department of Computer System Engineering, Arab American University, Jenin, Palestine [1, 2]

***Abstract*—The escalating challenge of waste management, particularly in developed nations, necessitates innovative approaches to enhance recycling and sorting efficiency. This study investigates the application of Convolutional Neural Networks (CNNs) for landfill waste classification, addressing the limitations of traditional sorting methods. We conducted a performance comparison of five prevalent CNN models—VGG-16, InceptionResNetV2, DenseNet121, Inception V3, and MobileNetV2—using the newly introduced "RealWaste" dataset, comprising 4,752 labeled images. Our findings reveal that EfficientNet achieved the highest average testing accuracy of 96.31%, significantly outperforming other models. The analysis also highlighted common challenges in accurately distinguishing between metal and plastic waste categories across all models. This research underscores the potential of deep learning techniques in automating waste classification processes, thereby contributing to more effective waste management strategies and promoting environmental sustainability.**

Keywords*—Waste management; deep learning; waste classification; real-waste dataset; performance comparison

I. Introduction

The increase in waste generation, particularly in developed countries, poses a significant challenge to effective waste management and recycling efforts. By 2050, it is projected that developed nations will experience a 19% increase in per capita daily waste production, emphasizing the critical need for more efficient waste management strategies [1]. Traditional waste sorting methods, such as manual sorting and visual inspection, have limitations in terms of subjectivity, scalability, and labor requirements [1]. To address these challenges, the integration of deep learning techniques, particularly Convolutional Neural Networks (CNNs), into waste sorting processes can enhance automation and improve waste classification based on its features [2], [3].

CNNs are a class of deep learning models that excel in processing visual data, making them well-suited for tasks like waste classification [3]. These networks automatically extract relevant information from input data through their layers, with convolutional layers specifically extracting spatial features from images, making CNNs highly efficient for image-related tasks [4]. By leveraging advanced technologies like deep learning, waste sorting processes can be optimized, recycling rates can be increased, and a more sustainable waste management system can be achieved [5].

Automated waste classification systems, powered by deep learning models like CNNs, have become essential for addressing the global waste problem and promoting sustainable development [6]. These systems offer a more objective and scalable approach to waste sorting compared to traditional methods, contributing to more efficient recycling processes and waste management overall [7]. The incorporation of deep learning methods in waste classification not only streamlines the sorting process but also plays a crucial role in achieving environmental sustainability by reducing waste and promoting recycling efforts [8].

In conclusion, the adoption of deep learning techniques, particularly CNNs, in waste classification is pivotal for enhancing automation, improving waste sorting accuracy, and ultimately aiming to create a waste management system that is more effective and sustainable in light of the rising amount of garbage produced worldwide.

This paper presents a performance comparison of five common CNN pre-trained models applied on a dataset called "RealWaste" and provides a critical analysis of the results to see the impact of the quality of the dataset and more detailed classes of the waste. Hence, the main contributions of this paper are as follows:

- Provide a performance comparison of five common CNN models in landfill waste classification.
- Evaluate the performance of the selected models when the type of material over different items is important to be detected.
- Employ transfer learning and fine-tune the learning process using the scheduling of the learning rate.
- Achieve superior classification accuracy compared to previous work.

The remainder of this paper is structured as follows: Section II discusses the related works. The RealWaste dataset and the proposed models for evaluation are detailed in Section III and Section IV. Section V discusses the results and outcomes, and Section VI concludes the paper.

II. Literature Review

Table I presents the top accuracies achieved by different datasets in waste classification tasks. It includes datasets such as RealWaste and DiversionNet, Waste dataset, Custom dataset, Sekar's waste classification, OrgalidWaste, Waste Classification data, and the proposed work using RealWaste. The accuracy values range from 49.69% to 99.43%, highlighting the varying performance levels of the datasets in accurately classifying waste materials.

TABLE I. COMPARATIVE ANALYSIS OF DATASETS AND TOP ACCURACIES IN WASTE CLASSIFICATION

Ref.	Dataset	Top Accuracy
[9]	RealWaste and DiversionNet	49.69% using DiversionNet and 89.19% using RealWaste
[10]	Waste dataset	70%
[11]	Custom dataset	99.43%
[12]	Sekar's waste classification	80.88%
[13]	OrgalidWaste	88.42%
[14]	Waste Classification data	96.7%
Our proposed work	RealWaste	96.31%

In study [9], the authors propose a new dataset called RealWaste and evaluate the performance of five deep learning models (VGG-16, InceptionResNetV2, DenseNet121, Inception V3, and MobileNetV2) for waste classification using the "RealWaste" dataset and the existing "DiversionNet" dataset. The classification accuracy for the "DiversionNet" dataset was limited to 49.69% for the "RealWaste" dataset, the models were able to achieve much higher classification accuracy where Inception V3, reached 89.19% classification accuracy on the full spectrum of labels required for accurate waste modeling.

In research [10], the authors use a custom CNN model architecture with four and five convolution layers to classify four categories of solid waste including plastic, glass, organic, and paper materials. They use a "Waste dataset" that contains 100 RGB images for each category. The five-layer DCNN architecture achieved a 70% accuracy rate in distinguishing the different waste types, while the four-layer architecture had a 61.67% accuracy rate. Plastic waste was the most challenging to classify accurately, with 37% and 56.7% accuracy rates in the four-layer and five-layer architectures, respectively. The key limitations were around classification accuracy, particularly for plastic waste, as well as the need for further optimization and exploration of a broader range of waste types and real-world application considerations.

In study [11], the authors use a custom CNN model with two, six, and eight convolutional layers with a custom dataset of 878 carrot images captured in-house, which they preprocessed and augmented to train and evaluate their proposed CNN models. The authors found that the eight convolutional layers model with the mixed pooling layer achieved the best performance, reaching 99.43% accuracy in classifying regular and irregular carrot shapes on 24x24 pixel images. The study was conducted using a dataset of 878 carrot images, which may be considered a relatively small dataset for training deep learning models. Also, the study was conducted in a controlled laboratory setting using images captured under specific lighting conditions. The authors acknowledge that while the proposed CNN-based approach showed promising results in the laboratory setting, further research and real-world validation would be needed to fully assess the practical applicability and limitations of the system.

In study [12], the authors developed a bespoke 5-layer CNN architecture and trained it on two different image resolutions (80x45 and 225x264 pixels) of the augmented "Sekar's waste classification" dataset consisting of 25,077 images of organic (13,966) and recyclable (11,111) waste objects. The research aims to explore the possibility of training an efficient lightweight model with high accuracy and less computational demand compared to standard CNN architectures. The smaller model (80.88%) outperformed the larger model (76.19%), but the larger model seems more generalizable based on the observed behavior of loss and accuracy during training, validation, and testing. The key limitations include the data paucity and limited categories.

In study [13], the authors use various CNN models for waste classification, including AlexNet, GoogLeNet, EfficientNet-B0, and ResNet-50 with a transfer learning approach. Also, they use a custom four-layer CNN. They used a dataset called "OrgalidWaste" which comprises approximately 5,600 images categorized into four waste classes: organic, glass, metal, and plastic. The performance evaluation of the models in the study was based on accuracy and cross-entropy loss observed during training, validation, and testing. The VGG16 model achieved the highest accuracy of 88.42%, outperforming other CNN architectures like VGG19, Inception-V3, and ResNet50. The study highlighted the need to further enhance classification accuracy for practical deployment despite the promising results obtained.

In study [14], the authors use convolutional neural networks (CNNs) and faster region-based convolutional neural networks (R-CNNs) to classify e-waste. The proposed system aims to facilitate communication between individuals requesting WEEE pickup and waste collection companies, enabling efficient collection planning based on the identified type and size of the e-waste items. They used a dataset called "Waste Classification data" containing 24,705 images of refrigerators, washing machines, and monitors/TVs. The geometric transformations were used as data augmentation with 13 transformations such as rotation, color transformation, zoom, and blur. The proposed CNN model achieved an average accuracy of 90-96.7% and the faster R-CNN network provided slightly lower accuracy, around 90% on average, but had the advantage of being able to detect and determine the size of the objects in the images. The key limitations were around the limited e-waste categories.

III. DATASET

The study leveraged the newly developed landfill waste dataset, "RealWaste," which comprises 4,752 raw and fully labeled RGB images. This dataset was meticulously collected during the biannual residential waste audit at the Wollongong Waste and Resource Recovery Centre's landfill [9]. The collection process involved capturing images of various waste items as they were sorted, ensuring a representative sample of the types of materials commonly found in residential waste.

Fig. 1 shows samples of RealWaste images with their label.

(a) (b) (c) (d)

Fig. 1. Sample images from the dataset for: (a) Cardboard; (b) Food Organics; (c) Glass; (d) Metal.

The RealWaste dataset includes samples across nine distinct labels representing different landfill waste categories including Cardboard, Food Organics, Glass, Metal, Miscellaneous Trash, Paper, Plastic, Textile Trash and Vegetation.

An analysis of the dataset reveals that it exhibits some degree of imbalance in the distribution of images across the labels. Table II presents the count and percentage of images for each label, highlighting the variations.

TABLE II. LABELS AND IMAGE COUNT IN DATASET

Label	Images Count	Percentage Of Each Label in The Dataset
Cardboard	461	9.69%
Food Organics	411	8.65%
Glass	420	8.83%
Metal	790	16.62%
Miscellaneous Trash	495	10.41%
Paper	500	10.51%
Plastic	921	19.38%
Textile Trash	318	6.69%
Vegetation	436	9.17%

This imbalance may pose challenges for model training, particularly in ensuring that the model generalizes well across all categories. The predominance of plastic images, for instance, could lead to bias in classification outcomes if not adequately addressed.

IV. METHODOLOGY

This section addresses a significant gap in waste classification literature, focusing on dataset limitations and inadequate labeling in waste auditing studies. To enhance the accuracy and practicality of waste classification models, data is preprocessed and augmented for use by the pre-trained CNN models EfficientNet, GoogLeNet, ResNet-152, ShuffleNet, and VGG-19. The objectives include evaluating the use of clean material datasets for training models in real waste classification, comparing dataset approaches by training on real waste samples, and identifying the best model for waste classification in real-world scenarios.

As shown in Fig. 2, the proposed methodology consists of the following main steps:

- Data Splitting: The dataset was split into training, validation, and testing subsets using a standard split ratio, where 50% of the data was used for training, 20% for validation, and 30% for testing. This ensures that the model is trained on a diverse set of data, validated on a separate subset, and finally tested on unseen data [15].

- Hyperparameter tuning: Hyperparameters–such as batch size, learning rate, number of epochs, momentum, and learning rate decay–are set. These hyperparameters play a crucial role in the training process and should be carefully chosen through experimentation and validation [16].

- Training and Validation: Each model is trained and validated using a stochastic gradient descent optimizer. Additionally, the use of automatic mixed precision (auto-cast) is considered to accelerate training without sacrificing model accuracy. Furthermore, a learning rate scheduler is employed to adjust the learning rate during training [17].

- Model Evaluation: Each model uses metrics such as accuracy, precision, recall, F1 score, receiver operating characteristic curve, and confusion matrix analysis. These metrics provide a comprehensive understanding of the model's performance, such as correctly classifying instances, handling imbalanced classes, and discriminating between classes [18].

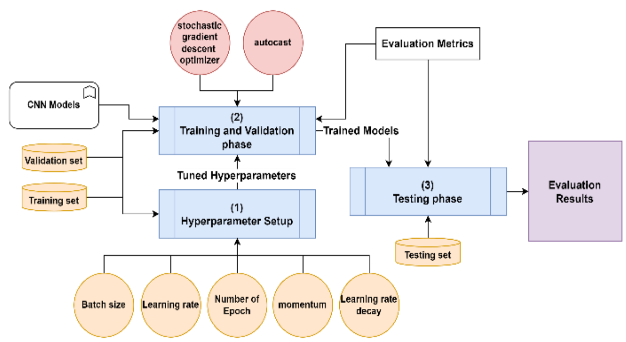


Fig. 2. The proposed methodology.

A. Data Preprocessing and Augmentation

CNNs typically require input images to be of a fixed size to avoid issues with model training and performance. Variations in image dimensions can be addressed by resizing images before inputting them into the CNN. This can be done using techniques like Bicubic interpolation, which involves averaging 16 neighboring pixels to determine pixel values in the resized image [19], [20]. Data augmentation techniques, which improve models' ability to generalize and perform well on diverse datasets and solve unbalanced dataset issues such as:

- Color Jitter which introduces random variations in the hue and saturation levels of images to augment the dataset. The hue indicates the range of random hue adjustments applied to the image, while saturation represents the range of random saturation adjustments.

- Randomly rotates images within the specified range of degrees.

The incorporation of augmented images with diverse textures and appearances enhanced the diversity and richness of the dataset. As a result, this contributed to an improved generalization of models and enhanced their performance across all classes. Each image in the dataset was used to generate two new images: The first was generated by adding Color Jitter with a hue value of 0.05 and a saturation of 0.05 to the original image. The second was generated by applying random rotations to images in the dataset with a range from 0 to 180 degrees.

B. The Selected Models

The CNN models including EfficientNet, GoogLeNet, ResNet-152, ShuffleNet, and VGG-19 were selected for this experimental study to analyze the performance in classifying landfill waste due to their proven capabilities in image classification tasks. EfficientNet is renowned for its efficiency in balancing model size and accuracy. GoogLeNet introduces the inception module for feature extraction. ResNet-152 utilizes residual connections to tackle the vanishing gradient problem. ShuffleNet emphasizes computational efficiency through channel shuffling. VGG-19 is acknowledged for its deep architecture with small convolutional filters [21], [22], [23]. The strengths of these models in image classification position them as ideal candidates for accurately classifying landfill waste:

- EfficientNetV2, introduced in 2021 by Tan and Le, improves training speed and parameter efficiency compared to the original EfficientNet. It addresses slow training with larger image resolutions by combining MBConv and Fused-MBConv blocks through neural architecture search and model scaling. This optimization enhances training efficiency [24].
- GoogLeNet: GoogLeNet developed by Google in 2015, is a deep convolutional neural network for image classification. It uses multiple convolutional layers with different filter sizes and pooling operations to extract features at various scales. The architecture consists of 19 layers, featuring inception modules for feature extraction, auxiliary classifiers to address the vanishing gradient issue and overfitting, and ensuring computational efficiency. It also includes max-pooling layers, an average pooling layer, a dropout layer, and a linear layer for the final output [25].
- ResNet-152: ResNet-152 was introduced in 2015 by Microsoft. It is part of the ResNet (short for Residual Network) family of models, known for their deep structure and the use of residual connections. ResNet-152 specifically has 152 layers, making it a very deep network, and it has been widely used for various computer vision tasks, such as image classification and object detection [26].
- ShuffleNet V2: Introduced in 2018 as an evolution of the original ShuffleNet, focuses on enhancing computational efficiency and model performance. By integrating channel shuffling and pointwise group convolutions, it improves performance while preserving computational efficiency. These elements enhance its effectiveness in feature extraction and representation learning. With 164 layers, ShuffleNet V2 incorporates operations like depthwise separable convolutions, concatenation, and channel shuffling to facilitate efficient information exchange and feature extraction within the network [27], [28].
- VGG19: VGG19, a deep convolutional neural network that comprises 19 layers, was developed in 2014. It is known for its simplicity and effectiveness in image recognition tasks. The network architecture consists of a series of convolutional layers that are followed by max pooling layers and culminates in three fully connected layers [29].

C. Hyperparameters and Optimization Technique

Hyperparameters are parameters that govern the learning process and dictate the values of the model parameters acquired by a learning algorithm. The use of the prefix "hyper" indicates the significance of the parameters in determining both the learning process and resulting model parameters [30]. Specifically, hyperparameters are settings or configurations that are not learned from the data but are set before training the model, and in the proposed methodology, the following hyperparameters were used: learning rate, loss function, number of epochs, and batch size.

Optimizers are essential for adjusting model weights and learning rates to minimize errors or maximize efficiency. For instance, stochastic gradient descent is a popular optimization algorithm in deep learning, a variation of gradient descent. It minimizes loss functions by training models. Unlike traditional gradient descent, stochastic gradient descent computes gradients using data subsets, known as "mini-batches," making it faster and more scalable for large datasets. This approach accelerates convergence, especially beneficial for handling extensive datasets [31], [32]; therefore, it was used in the proposed methodology. The objective function $f(x)$, which is the average loss function, is given by the following equation:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x) \tag{1}$$

where n is the number of input data taken from the training dataset. The gradient of the objective function is computed for each iteration of the training phase by the following equation:

$$\nabla f(x) = \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(x) \tag{2}$$

The stochastic gradient descent algorithm reduces the computational cost at each iteration by randomly shuffling the training data to ensure that the mini-batches used for computing the gradients are representative of the entire dataset and compute the gradient $\nabla f_i(x)$ to update x by the following equation [33]:

$$\mathrm{x} \leftarrow x - \eta \nabla f_i(x) \tag{3}$$

where η is the learning rate.

D. Evaluation Metrics

Various evaluation metrics were utilized to appraise the performance of the selected models, including the confusion matrix, accuracy, F1 score, precision, recall, and receiver operating characteristic (ROC) curve [34]. The confusion matrix evaluates the classification performance of the CNN model by juxtaposing actual and predicted values. In this matrix, rows correspond to actual values while columns represent predicted values. The results obtained from this evaluation encompass four potential outcomes: True Positive (TP) - denoting correct prediction of the positive class, False Positive (FP) - indicating incorrect prediction of the positive class, True Negative (TN) - signifying accurate prediction of the negative class, and False Negative (FN) - representing erroneous prediction of the negative class [35].

Accuracy refers to the extent to which the model effectively categorizes all instances within a dataset. It is computed by dividing the total number of correct predictions by the overall number of predictions made [36]. The following equation calculates accuracy:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{4}$$

The F1 score is a numerical measure of the balance between precision and recall [37]. It is calculated by taking the harmonic mean of precision and recall using the following equation:

$$F1\, score = 2 * \frac{(precision * recall)}{(precision + recall)} \tag{5}$$

Where precision is the ratio of true positives to the sum of true positives and false positives, measuring the accuracy of positive predictions. Recall, on the other hand, is the proportion of true positives to the sum of true positives and false negatives, indicating the ability to identify actual positives accurately.

The ROC curve evaluates a CNN model's ability to distinguish labels by examining the true positive rate (TPR) and false positive rate (FPR) at different thresholds. It assesses the model's accuracy in classification by graphing TPR against FPR at various threshold levels [38]. The TPR and FPR are calculated using the following equations:

$$TPR = \frac{TP}{(TP + FN)} \tag{6}$$

$$FPR = \frac{FP}{(FP + FN)} \tag{7}$$

V. Results and Discussions

A. Experimental Setup

The original dataset was cleaned, preprocessed, and enhanced as outlined in Section III to prepare it for training CNN models. The image quantity was increased to 19,008 labeled images, with each label representing around 3% of the dataset, totaling 9 unique labels. One-hot encoding was utilized for label encoding. Various hyperparameters such as learning rate, loss function, epochs, batch size, optimizer, momentum, learning rate decay, and weight initialization were set for training the CNN models as shown in Table III. The selected models' weights were initialized by loading pre-trained weights from ImageNet, enhancing their performance and efficiency while reducing the need for extensive training on the dataset.

TABLE III. Hyperparameters and their Values Tuning

Hyperparameters	Value
Learning rate	0.001
Loss function	Cross-entropy loss
Number of epochs	50
Batch sizes	64
Optimizer	Stochastic gradient descent
Momentum	0.9
Learning rate decay	0.0005
Weight initialization	Transfer learning

All experiments were carried out locally on a computer with the following specifications:

- Processor: AMD Ryzen 9 5900HX, 3301 MHz, 8 Core(s), 16 Logical Processor(s).
- Physical Memory (RAM): 32.0 GB.
- Graphics Card: NVIDIA GeForce RTX 3080 Laptop GPU.

B. Models Training Performance

All models were trained using the same dataset and with hyperparameter values as listed in Table III to analyze the learning behavior of each model. Table IV presents the average training and validation accuracy, as well as the average training and validation loss for various models including EfficientNet, GoogLeNet, ResNet-152, ShuffleNet, and VGG-19. ResNet-152 has the highest average training among the models listed, with a value of 99.07%.

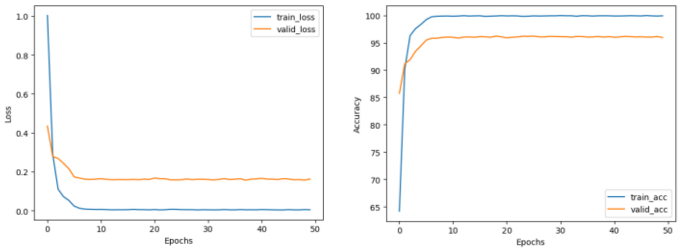
TABLE IV. Hyperparameters and their Values Tuning

#	Model	Average Training Accuracy	Average Validation Accuracy	Average Training Loss	Average Validation Loss
1	EfficientNet	98.82%	95.57%	0.036	0.174
2	GoogLeNet	98.93%	92.93%	0.035	0.234
3	ResNet-152	99.07%	94.63%	0.029	0.198
4	ShuffleNet	92.83%	86.42%	0.270	0.430
5	VGG-19	97.94%	91.22%	0.063	0.343

ShuffleNet has the lowest average training accuracy at 92.83% but still demonstrates strong performance. EfficientNet and GoogLeNet demonstrate competitive performance in terms of accuracy and loss, while VGG-19 shows slightly lower accuracy, but relatively lower loss compared to ResNet-152.

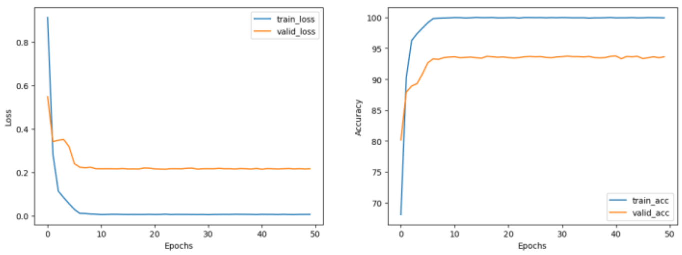
Overall, the analysis shows that ResNet-152 performs exceptionally well in terms of both accuracy and loss, while ShuffleNet appears to face challenges in generalizing to validate data. This comparative analysis can provide valuable insights for selecting the most suitable model based on specific performance criteria.

Fig. 3 to Fig. 7 visualize the loss values and average accuracy percentage for all models per epoch.



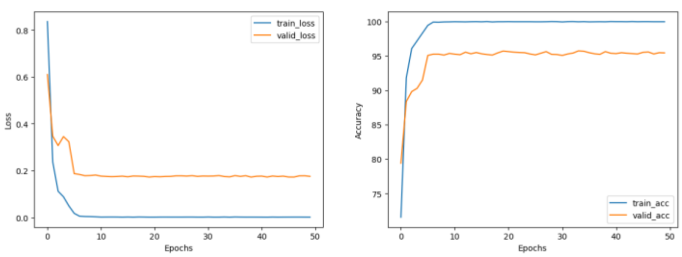
(a) (b)

Fig. 3. EfficientNet training performance: (a) Losses values per epoch (b) Average accuracy percentage values per epoch.



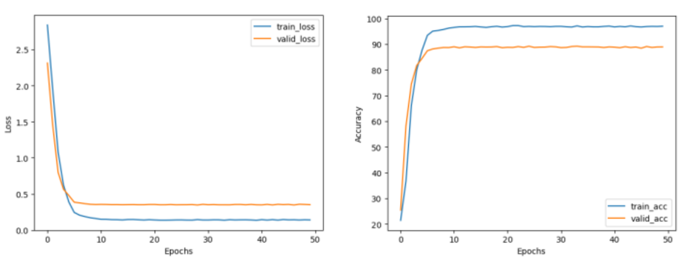
(a) (b)

Fig. 4. GoogLeNet training performance: (a) Losses values per epoch (b) Average accuracy percentage values per epoch.



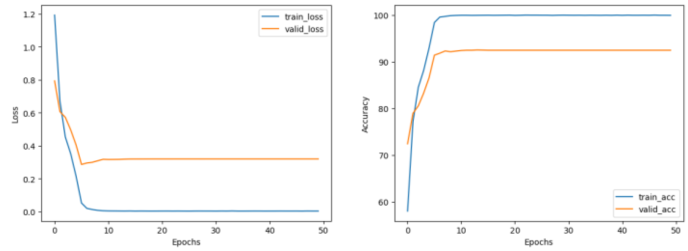
(a) (b)

Fig. 5. ResNet-152 training performance: (a) Losses values per epoch (b) Average accuracy percentage values per epoch.



(a) (b)

Fig. 6. ShuffleNet training performance: (a) Losses values per epoch (b) Average accuracy percentage values per epoch.



(a) (b)

Fig. 7. VGG-19 training performance: (a) Losses values per epoch (b) Average accuracy percentage values per epoch.

C. Models Testing Performance

Table V provides a comprehensive overview of the performance metrics including accuracy, precision, recall, and F1 score for different models. Among these models, EfficientNet stands out with the highest Average Testing Accuracy of 96.31%, highlighting its effectiveness in classification tasks. This exceptional performance can be attributed to Efficient Net's sophisticated architecture that incorporates compound scaling and efficient model scaling methods, allowing it to achieve superior accuracy.

TABLE V. HYPERPARAMETERS AND THEIR VALUES TUNING

#	Model	Average Testing Accuracy	Average Precision	Average Recall	Average F1 Score
1	EfficientNet	96.31%	0.9342	0.9631	0.9643
2	GoogLeNet	94.25%	0.8965	0.9425	0.9432
3	ResNet-152	95.49%	0.9183	0.9549	0.9555
4	ShuffleNet	89.11%	0.8113	0.8911	0.8924
5	VGG-19	92.91%	0.8727	0.9291	0.9296

ResNet-152 and GoogLeNet also demonstrate strong performance with accuracy rates of 95.49% and 94.25% respectively. ResNet-152, known for its deep architecture with residual connections, excels in capturing intricate features within the data, leading to high precision, recall, and F1 scores. On the other hand, Google Net's inception modules and efficient use of parameters contribute to its competitive performance across all metrics.

ShuffleNet and VGG-19, while slightly lower in accuracy compared to the top performers, still exhibit respectable results. Shuffle Net's emphasis on computational efficiency through channel shuffle operations enables it to achieve a balance between accuracy and resource utilization. VGG-19, with its deeper architecture comprising multiple convolutional layers, maintains a strong performance across precision, recall, and F1 score metrics.

D. Models Performance Analysis

To gain a deeper understanding of the models' classification performance, a detailed analysis was conducted using the AUC, as shown in Fig. 8.

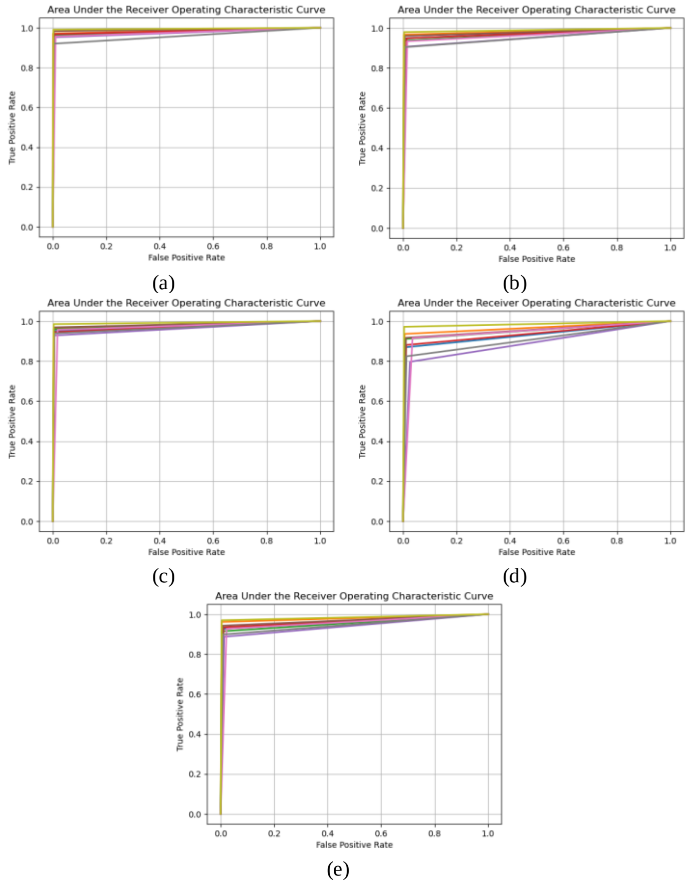


Fig. 8. AUC values for different Labels across various models including (a) EfficientNet (b) GoogLeNet (c) ResNet-152 (d) ShuffleNet and (e) VGG-19.

EfficientNet demonstrates strong performance across most classes with consistently high AUC values, particularly excelling in classes such as Food Organics and Vegetation where it achieved AUC scores of 0.99. This suggests that EfficientNet is effective in distinguishing these classes from others with high accuracy. GoogLeNet also performs well overall, with notable AUC scores for classes such as Glass and Vegetation. However, it shows slightly lower performance compared to EfficientNet in some classes like Metal and Miscellaneous Trash. ResNet-152 showcases consistent performance across various classes, with competitive AUC values for most categories. It performs particularly well in distinguishing classes like Cardboard and Paper. ShuffleNet exhibits varying performance across different classes, with lower AUC scores for categories such as Miscellaneous Trash and Textile Trash compared to other models. VGG-19, similar to GoogLeNet, demonstrates strong performance in classes like Glass and Vegetation but shows lower AUC values for categories like Miscellaneous Trash and Textile Trash.

In summary, EfficientNet stands out as a top performer in this analysis, followed closely by ResNet-152 and GoogLeNet. These models show effectiveness in classifying diverse classes accurately, with variations in performance observed across different models and classes.

Also, the confusion matrix was utilized to analyze the model's classification performance in correctly classifying different waste categories and reveal the presence of false negatives and false positives. Fig. 9 to Fig. 13 visualize the confusion matrix for the five models.

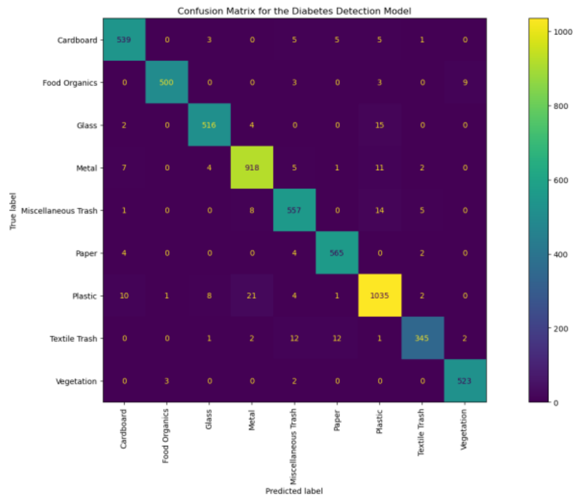


Fig. 9. Confusion Matrix for EfficientNet model in classifying landfill waste.

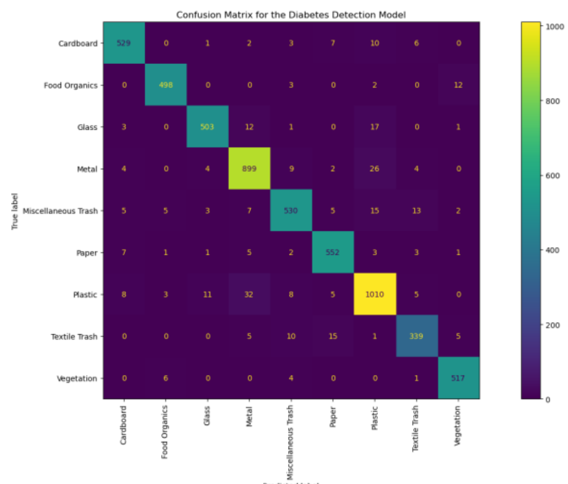


Fig. 10. Confusion Matrix for GoogLeNet model in classifying landfill waste.

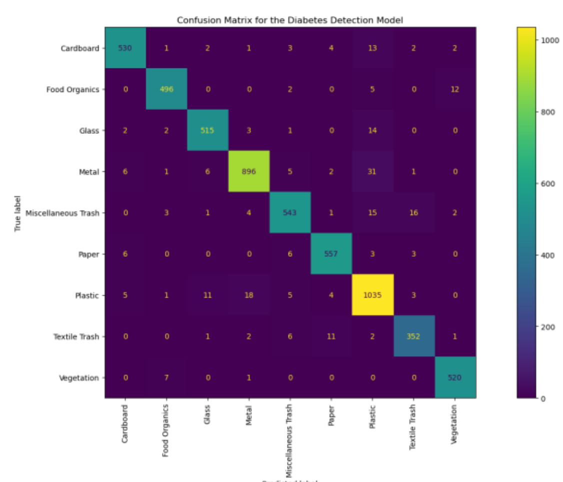


Fig. 11. Confusion Matrix for ResNet-152 model in classifying landfill waste.

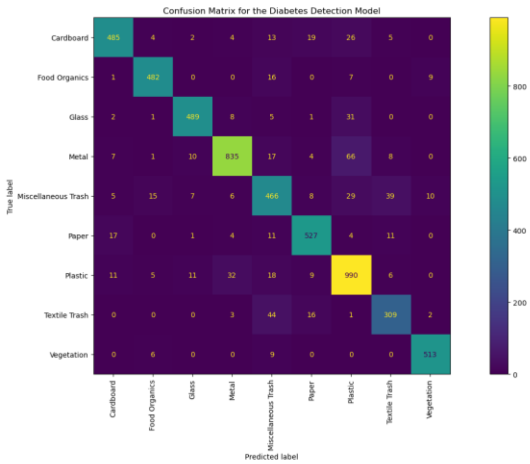


Fig. 12. Confusion Matrix for ShuffleNet model in classifying landfill waste.

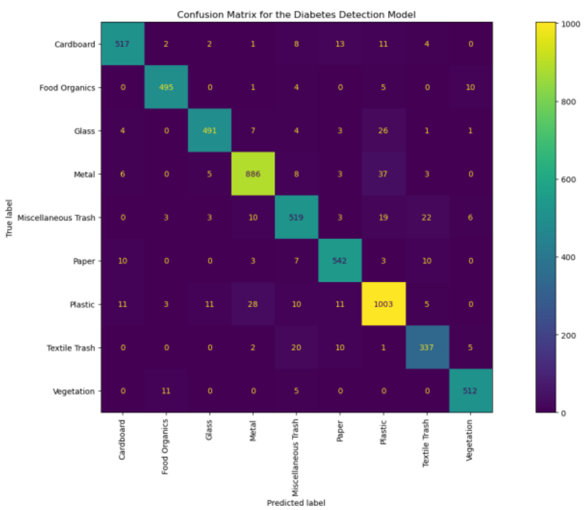


Fig. 13. Confusion Matrix for VGG-19 model in classifying landfill waste.

The confusion matrix for all models shows that the model's performance reveals notable confusion among various waste labels, particularly between metal and plastic categories. For EfficientNet, the model achieved a high accuracy rate of 92.41% in correctly identifying metal objects but misclassified four items as plastic. The accuracy of the plastic label was slightly lower at 84.78%, with six items mistakenly classified as metal and one item as cardboard. For GoogLeNet, the model's performance metrics show a similar trend with notable confusion between metal and plastic categories. Similarly, the ResNet-152 model exhibited challenges in distinguishing between metal and plastic categories, leading to misclassification. VGG-19 model's performance also indicated confusion between metal and plastic labels, impacting the accuracy of classification. In summary, all models struggled with distinguishing between metal and plastic waste categories, highlighting a common challenge across the different models in accurately classifying these materials. The difficulty in distinguishing between metal and plastic waste categories in the models could be attributed to several factors:

- Multi-Structured Shapes and Textures: Both metal and plastic items have varied shapes and textures, which further complicates the classification problem for the models.
- Data Variability: The reason why models fail to differentiate between waste materials made of metal and those that are made of plastic may partly be attributed to the lack of diversity in examples used during training concerning these two classes.

To address the challenges faced by models in accurately classifying metal and plastic waste, several strategies can be implemented:

- **Diverse Data Sources**: Where the current dataset is established, further research could explore the utilization of other sources of images taken in different surroundings and conditions of light. This would help create a more comprehensive dataset that captures a wider range of metal and plastic items.
- **Improved Image Augmentation**: While augmentation has been done, checking out advanced augmentation techniques, such as changes in brightness, contrast, and adding artificial noise, might yield a better generalization performance across different conditions.

E. Comparison with State-of-the-Art

To ensure an unbiased comparison, we have chosen to evaluate our work by benchmarking it against other studies that have utilized the same dataset. This approach allows for a fair assessment of the effectiveness and performance of our methodology within the context of the specific dataset, promoting a more accurate evaluation of our contributions in the field.

In study [9], they achieved an accuracy of 49.69% using DiversionNet and 89.19% using RealWaste. The accuracy of using RealWaste was notably higher compared to DiversionNet, indicating the effectiveness of RealWaste in the classification process. However, our work demonstrated a higher accuracy of 96.31% using the RealWaste dataset. It showed a significant improvement in accuracy compared to Work 1's results with RealWaste.

In summary, our proposed work exhibited superior performance in waste classification using the RealWaste dataset compared to study [9], showcasing advancements in accuracy and potentially innovative approaches in the classification process.

VI. Conclusion

Incorporating deep learning techniques, especially Convolutional Neural Networks (CNNs), into waste classification processes is crucial for enhancing automation, improving waste sorting accuracy, and striving towards a more sustainable and efficient waste management system amidst the escalating global waste production. The performance evaluation of five common CNN pre-trained models on the RealWaste dataset not only demonstrated advancements in accuracy but also highlighted potential innovative approaches in waste classification methodologies. This study contributes to the continuous enhancement of waste management strategies and the promotion of environmental sustainability through the utilization of cutting-edge technologies like deep learning. Notably, EfficientNet emerged as the top performer with the highest Average Testing Accuracy of 96.31%, underscoring its effectiveness in classification tasks. Additionally, the performance of five well-known CNN pre-trained models was compared in this research using the RealWaste dataset as a test. The evaluation aimed to understand the impact of the dataset quality and the inclusion of more detailed waste classes, shedding light on the importance of data quality in achieving accurate waste classification outcomes. By addressing these aspects, the research contributes to advancing waste management practices and fostering environmental sustainability through advanced deep-learning methodologies. Looking ahead, future work will focus on expanding the dataset to include more diverse samples and exploring advanced data augmentation techniques to address class imbalances. These efforts aim to further enhance model robustness and improve waste classification systems, ultimately contributing to more effective waste management practices and fostering environmental sustainability.

REFERENCES

[1] A. Sevinç and F. Ozyurt, "Classification of recyclable waste using deep learning architectures," FIRAT UNIVERSITY JOURNAL OF EXPERIMENTAL AND COMPUTATIONAL ENGINEERING, vol. 1, no. 3, 2022, doi: 10.5505/fujece.2022.83997.

[2] A. U. Gondal et al., "Real time multipurpose smart waste classification model for efficient recycling in smart cities using multilayer convolutional neural network and perceptron," Sensors, vol. 21, no. 14, 2021, doi: 10.3390/s21144916.

[3] V. Gadre, S. Sashte, and A. Sarnaik, "WASTE CLASSIFICATION USING RESNET-152," INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, vol. 07, no. 01, 2023, doi: 10.55041/ijsrem17421.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun ACM, vol. 60, no. 6, 2017, doi: 10.1145/3065386.

[5] D. O. Melinte, A. M. Travediu, and D. N. Dumitriu, "Deep convolutional neural networks object detector for real-time waste identification," Applied Sciences (Switzerland), vol. 10, no. 20, 2020, doi: 10.3390/app10207301.

[6] Q. Zhang et al., "Recyclable waste image recognition based on deep learning," Resources, Conservation and Recycling, vol. 171. 2021. doi: 10.1016/j.resconrec.2021.105636.

[7] Sivaranjani S, Priyanka Sherllyn S, Deepaharshini G R, and Eunice J, "Green Symphony: A Brief Review of Waste Segregation Techniques," International Research Journal on Advanced Engineering Hub (IRJAEH), vol. 2, no. 03, 2024, doi: 10.47392/irjaeh.2024.0056.

[8] K. O. Mohammed Aarif, C. Mohamed Yousuff, B. A. Mohammed Hashim, C. Mohamed Hashim, and P. Sivakumar, "Smart bin: Waste segregation system using deep learning-Internet of Things for sustainable smart cities," Concurr Comput, vol. 34, no. 28, 2022, doi: 10.1002/cpe.7378.

[9] S. Single, S. Iranmanesh, and R. Raad, "RealWaste: A Novel Real-Life Data Set for Landfill Waste Classification Using Deep Learning," Information (Switzerland), vol. 14, no. 12, 2023, doi: 10.3390/info14120633.

[10] A. Altikat, A. Gulbe, and S. Altikat, "Intelligent solid waste classification using deep convolutional neural networks," International Journal of Environmental Science and Technology, vol. 19, no. 3, 2022, doi: 10.1007/s13762-021-03179-4.

[11] A. Jahanbakhshi, M. Momeny, M. Mahmoudi, and P. Radeva, "Waste management using an automatic sorting system for carrot fruit based on image processing technique and improved deep neural networks," Energy Reports, vol. 7, 2021, doi: 10.1016/j.egyr.2021.08.028.

[12] R. Faria, F. Ahmed, A. Das, and A. Dey, "Classification of Organic and Solid Waste Using Deep Convolutional Neural Networks," in IEEE Region 10 Humanitarian Technology Conference, R10-HTC, 2021. doi: 10.1109/R10-HTC53172.2021.9641560.

[13] N. Nnamoko, J. Barrowclough, and J. Procter, "Solid Waste Image Classification Using Deep Convolutional Neural Network," Infrastructures (Basel), vol. 7, no. 4, 2022, doi: 10.3390/infrastructures7040047.

[14] P. Nowakowski and T. Pamuła, "Application of deep learning object classifier to improve e-waste collection planning," Waste Management, vol. 109, 2020, doi: 10.1016/j.wasman.2020.04.041.

[15] K. M. Kahloot and P. Ekler, "Algorithmic Splitting: A Method for Dataset Preparation," IEEE Access, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3110745.

[16] P. Probst, A. L. Boulesteix, and B. Bischl, "Tunability: Importance of hyperparameters of machine learning algorithms," Journal of Machine Learning Research, vol. 20, 2019.

[17] K. Wang, Y. Dou, T. Sun, P. Qiao, and D. Wen, "An automatic learning rate decay strategy for stochastic gradient descent optimization methods in neural networks," International Journal of Intelligent Systems, vol. 37, no. 10, 2022, doi: 10.1002/int.22883.

[18] A. Tharwat, "Classification assessment methods," Applied Computing and Informatics, vol. 17, no. 1, 2018, doi: 10.1016/j.aci.2018.08.003.

[19] P. Sengupta, A. Mehta, and P. S. Rana, "Enhancing Performance of Deep Learning Models with a Novel Data Augmentation Approach," in 2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023, 2023. doi: 10.1109/ICCCNT56998.2023.10308298.

[20] W. Liang, Y. Liang, and J. Jia, "MiAMix: Enhancing Image Classification through a Multi-Stage Augmented Mixed Sample Data Augmentation Method," Processes, vol. 11, no. 12, 2023, doi: 10.3390/pr11123284.

[21] M. H. Huynh, P. T. Pham-Hoai, A. K. Tran, and T. D. Nguyen, "Automated Waste Sorting Using Convolutional Neural Network," in Proceedings - 2020 7th NAFOSTED Conference on Information and Computer Science, NICS 2020, 2020. doi: 10.1109/NICS51282.2020.9335897.

[22] J. Bobulski and M. Kubanek, "Deep Learning for Plastic Waste Classification System," Applied Computational Intelligence and Soft Computing, vol. 2021, 2021, doi: 10.1155/2021/6626948.

[23] G. Lu, Y. Bin Wang, H. X. Xu, H. Y. Yang, and J. Zou, "Deep multimodal learning for municipal solid waste sorting," Sci China Technol Sci, vol. 65, no. 2, 2022, doi: 10.1007/s11431-021-1927-9.

[24] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," in Proceedings of Machine Learning Research, 2021.

[25] P. Aswathy, Siddhartha, and D. Mishra, "Deep GoogLeNet Features for Visual Object Tracking," in 2018 13th International Conference on Industrial and Information Systems, ICIIS 2018 - Proceedings, 2018. doi: 10.1109/ICIINFS.2018.8721317.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016. doi: 10.1109/CVPR.2016.90.

[27] R. Doss, J. Ramakrishnan, S. Kavitha, S. Ramkumar, G. Charlyn Pushpa Latha, and K. Ramaswamy, "Classification of Silicon (Si) Wafer Material Defects in Semiconductor Choosers using a Deep Learning ShuffleNet-v2-CNN Model," Advances in Materials Science and Engineering, vol. 2022, 2022, doi: 10.1155/2022/1829792.

[28] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "Shufflenet V2: Practical guidelines for efficient cnn architecture design," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018. doi: 10.1007/978-3-030-01264-9_8.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.

[30] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," Lecture Notes in Computer Science (including

subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7700 LECTURE NO, 2012, doi: 10.1007/978-3-642-35289-8_26.

[31] H. Zhang, K. Hao, L. Gao, B. Wei, and X. Tang, "Optimizing Deep Neural Networks Through Neuroevolution With Stochastic Gradient Descent," IEEE Trans Cogn Dev Syst, vol. 15, no. 1, 2023, doi: 10.1109/TCDS.2022.3146327.

[32] Q. Qian, R. Jin, J. Yi, L. Zhang, and S. Zhu, "Efficient distance metric learning by adaptive sampling and mini-batch stochastic gradient descent (SGD)," Mach Learn, vol. 99, no. 3, 2015, doi: 10.1007/s10994-014-5456-x.

[33] Y. Zhou, M. Zhang, J. Zhu, R. Zheng, and Q. Wu, "A Randomized Block-Coordinate Adam online learning optimization algorithm," Neural Comput Appl, vol. 32, no. 16, 2020, doi: 10.1007/s00521-020-04718-9.

[34] M. Soomro, M. A. Farooq, and R. H. Raza, "Performance evaluation of advanced deep learning architectures for offline handwritten character recognition," in Proceedings - 2017 International Conference on Frontiers of Information Technology, FIT 2017, 2017. doi: 10.1109/FIT.2017.00071.

[35] D. Silwal, "Confusion Matrix, Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures," Linkedin.

[36] H. Palus and D. Bereska, "COLOUR REPRODUCTION ACCURACY OF VISION SYSTEMS," in Computer Vision and Graphics, 2006. doi: 10.1007/1-4020-4179-9_40.

[37] R. Joshi, "Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures - Exsilio Blog," 2016.

[38] C. Marzban, "The ROC curve and the area under it as performance measures," Weather Forecast, vol. 19, no. 6, 2004, doi: 10.1175/825.1.