Garbage Classification from Visual Footprints: Using Transfer Learning Strategy

Zheyuan Xu and Nasim Hajari
Department of Mathematics & Information Technology, Concordia University of Edmonton, Edmonton, Canada

Keywords: Garbage Classification, Machine Learning, Computer Vision, Pre-Trained Models, Transfer Learning.

Abstract: This study investigates the application of computer vision models based on deep learning, to improve waste sorting and promote environmental sustainability. The research evaluates the effectiveness of Convolutional Neural Networks (CNNs) and transfer learning techniques by comparing the performance of eleven pre-trained models in classifying household waste from images into eight distinct categories. Through the implementation of fine-tuning, learning rate scheduling, and overfitting prevention strategies, the study optimizes model performance. Remarkably, the ConvNeXtBase and EfficientNetV2L models achieved impressive accuracy rates of 99.00% and 98.64%, respectively, underscoring the potential of modern CNN architectures in waste classification tasks. Furthermore, a comparative analysis with recent studies reveals that the dataset's size, quality, and category diversity play crucial roles in determining model performance, with larger and more diverse datasets enabling superior generalization. The originality of this research lies in its comprehensive, side-by-side comparison of multiple pre-trained models on garbage classification application. This offers valuable insights into balancing knowledge retention and adaptation to new tasks. The findings underscore the significant potential of advanced neural network architectures in enhancing waste management and recycling practices.

1 INTRODUCTION

Managing waste effectively is a significant global challenge, with inefficient sorting processes leading to increased environmental pollution and negatively affecting recycling efforts (Satvilkar, 2018). Traditional methods, which often depend on manual sorting, are not only time-consuming but also prone to errors. This study was inspired by the growing need for smarter solutions in garbage classification and waste management, especially as the volume of waste continues to rise with improving economies and living standards. Recent advancements in computer vision and machine learning present promising solutions for enhancing the accuracy of waste sorting. In this research, we take a close look at how different pre-trained computer vision models, perform in classifying waste into eight specific categories. Our goal is to fill a noticeable gap in the existing research by offering a detailed comparison of these models, specifically for transfer learning in the context of garbage classification. The remaining of this paper is organized as follows: the review of the relevant literature is presented in Section 2. Section 3 explains methodology including data preparation and preprocessing and selecting and fine-tuning the pre-trained models. Section 4 discusses the implementation of transfer learning, evaluates the models' performances, and compares their effectiveness, ultimately identifying the top two models for this task. Finally, Section 5 wraps up with conclusions.

2 LITERATURE REVIEW

In recent years, the application of transfer learning in garbage classification has steadily drawn the interest of many researchers. Although, the early results are not quite satisfactory, the algorithms efficiencies have been improved recently (Wang, 2022; Endah et al., 2020). Wu et al. (Wu et al., 2022) proposed an improved VGG16 model for waste classification, addressing overfitting and computational inefficiency issues found in the traditional model. The improved VGG16 achieved a high accuracy of 96.21% on the test set, with fewer parameters and shorter training time. This enhancement demonstrates its potential for real-world waste classification. Additionally, Ma et al. (Ma et al., 2022) compared MobileNet and MobileNetV2, finding that MobileNetV2 was more ac-

Xu, Z. and Hajari, N.
Garbage Classification from Visual Footprints: Using Transfer Learning Strategy.
DOI: 10.5220/0013151200003912

In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 3: VISAPP*, pages 403-409
ISBN: 978-989-758-728-3; ISSN: 2184-4321

curate and resource-efficient, making it ideal for embedded waste sorting systems, achieving 98.7% accuracy with minimal resources. Mehedi et al. (Mehedi et al., 2023) considered three different CNN architectures: a custom 6-layer baseline CNN, VGG16, and MobileNetV2. Their results showed that VGG16 had the best performance in terms of accuracy of 96.00%. MobileNetV2 also performed well with an accuracy of 95.51%, while the benchmark CNN had an accuracy of 90.61%. Shukurov (Shukurov, 2023) used five pre-trained CNN architectures for fine-tuning, finding that models optimized with the Adam optimizer, particularly ResNeXt50, achieved high validation accuracy with minimal overfitting. Lou and Gou (Lou and Gou, 2023) introduced transfer learning and Efficient Channel Attention mechanism (ECA) to enhance the feature extraction of the model. They introduced ECAE-Net, a model based on EfficientNetV2-S, which achieved 96.8% accuracy with fewer parameters and FLOPs, demonstrating the potential of incorporating attention mechanisms and optimization techniques. The research by Goel et al. (Goel et al., 2023) focused on four well-known models: DenseNet-161, ResNet-50, InceptionV3, and EfficientNet-B7. All models, augmented with migration learning methods, achieved high test accuracies, with ResNet50 leading at 94.70%. Haque et al. (Haque et al., 2024) researched customized CNN models with pre-trained models such as DenseNet, ResNet and Xception. DenseNet169 emerged as the top performer with a 99.58% accuracy, outperforming the other models in terms of both training and validation accuracy.

3 METHODOLOGY

The workflow used in this project is shown in Figure 1.

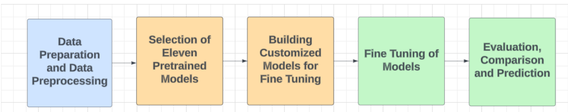

Figure 1: Workflow of the developed system.

The following subsections explain each phase in details.

3.1 Data Preparation and Preprocessing

3.1.1 Preparation of Datasets

Garbage dataset for this paper comes from the merging of two publicly available garbage datasets on the Kaggle website (Chang, 2018; Mohamed, 2021) The merged and modified dataset contains 6531 images divided into eight categories: batteries, brown glass, cardboard, green glass, metal, paper, plastic, and white glass. Table 1 shows the number of images contained in each category in the garbage dataset. Figure 2 shows an example of the final dataset.

Figure 2: An example of prepared dataset.

With batch_size = 32, img_height = 224, and img_width = 224, the garbage data is finally divided into training (70%), validation (15%) and test (15%) data, where the number of train batches is 143, the number of batches for both validation and test is 31.

3.1.2 Image Data Augmentation

Image data augmentation artificially increases the size of the image training dataset by generating many training instances from existing images. The generated instances should be as realistic as possible, that means, given an image from the augmented training dataset, a human should not be able to tell whether it was augmented or not (Géron, 2019).

We augment the data by flipping, rotating and adjusting is brightness randomly. This also helps the model to be more resilient towards variations in the position, orientation, and lighting conditions.

3.2 Transfer Learning

3.2.1 Reason to Use Transfer Learning

The pre-trained models are state-of-the-art models and generally trained on ImageNet database (Li et al.,

Table 1: Data distribution.

Classes	Number of Images	Classes	Number of Images
batteries	945	metal	769
brown glass	607	paper	1050
cardboard	891	plastic	865
green glass	629	white glass	775

2024). However, when these state-of-the-art models are applied to other related tasks/domain, they often suffer a considerable loss in performance due to their bias toward the training data and domain. Transfer learning uses the knowledge gained during training on a task and domain where sufficiently labeled data was available as a starting point (Ruder, 2017; Pan and Yang, 2009). In other words, transfer learning consists of taking features learned on one problem, and using them on a new, similar problem. Therefore, transfer leaning is particularly suitable to the tasks where the available dataset has too little data to train a full-scale model from scratch.

3.2.2 Techniques to Avoid Overfitting

Transfer learning could potentially lead to quick overfitting, therefore, avoiding overfitting is very important for transfer learning. In this paper, four techniques are used to avoid overfitting (Géron, 2019): Image data augmentation, Dropout, l2 norm regularization and Early stopping.

3.2.3 Building Customized Models for Fine Tuning

Eleven pre-trained models in Keras API (Keras-Team, 2024) are selected to implement this comparison study. Here, we take ConvNeXtBase model as example to show how the customized models used in this paper are built. For the new classification task with 8 categories, the top layer of ConvNeXtBase, originally designed for 1000 categories, is replaced with a new layer tailored to this task. The base model of ConvNeXtBase is combined with additional layers such as GlobalAveragePooling2D, Dropout, and Dense to build the final customized model. Similar modifications were applied to the other models in the study.

4 IMPLEMENTATION

4.1 Learning Rate Scheduling and Early Stopping

Choosing a good learning rate is very important. With a low learning rate, the training will eventually converge to the optimum, but it will last a very long time to finish. On the other hand, a high learning rate will make the training process very quickly at first, but it will end up jittering around the optimum and never really settling down. Generally, starting with a large learning rate and then reducing it gradually during the training yield a good solution faster than with the optimal constant learning rate (Géron, 2019). In this paper we use InverseTimeDecay schedule and set the initial learning rate to 0.00003. We also adapt an early stopping strategy to avoid overfitting, and eliminate long and unnecessary training times.

4.2 Fine-Tuning Techniques

The knowledge gained during training is stored in the weights of a neural network. However, the knowledge is not stored uniformly in all layers. In a stacked CNN for image classification, the initial convolution layers are only trained for extracting low-level features such as lines and edges, those complex feature such as animals and a body part or faces, are extracted by the higher convolution layers (Melcher and Silipo, 2020). Based on the extracted features, the last layer or output layer is responsible for classify the images.

Transfer learning in this paper uses the pre-trained models to build the transferred models with the weights gained from the source task as the starting point for the new task, which is garbage classification of eight categories. A key technique in transfer learning is fine tuning, which consists of unfreezing the entire base model obtained above or part of it, and re-training it on the new data with a very low learning rate. This can potentially achieve meaningful improvements, by incrementally adapting the pretrained features to the new data.

However, unfreezing the entire base model will increase training time, more importantly, it will take a risk to unlearn the knowledge gained from the source task. Therefore, we adapt a small learning rate to keep the learned knowledge from the source tasks as much as possible.

Table 2 shows the effect of freezing different layers (0%, 25%, 50%, 75%, and 100% of the layers from the lower layer to the upper layer) in the base model, respectively.

This experiment suggests that:

Table 2: Comparison of freezing different layers.

Freezing the layers in different percentages	Early stopping at epochs	Accuracy on test dataset
0%	143	95.08%
25%	149	95.41%
25%	95	95.89%
50%	54	95.13%
75%	65	95.15%
100%	227	91.4%

- Freezing 100% layers in the base model produces the lowest accuracy, 91.40%. This means only the weights in the classification head involve in training, limiting the learning ability of the model. Obviously, this is not a good choice.
- In the cases of freezing the layers in percentages: 0%, 25%, 50%, and 75% in the base model, there is a very little difference in accuracy, and all models obtain accuracies above 95%, specifically from 95.08% to 95.89%. Although the optimization path for each training process can be different, which leads to different training time. However, the required training time usually reduces as the freezing layers increase from 0% to 50%.
- Note that there is an obvious difference in training time for two cases of freezing 25% layers. This can be understood that different optimization paths require different training time.
- It is found that freezing 50% layers is a good balance between performance (accuracy) and training efficiency (number of iterations).

Based on the above analysis, this research takes into account a tradeoff between keeping the learned knowledge from the source task as much as possible and making the new model to adapt the new task; about 50% layers of the base model in the new model are frozen and at the same time, a very low learning rate is used for the fine tuning. In addition, each block needs to be all turned on or off. This is because the architecture includes a shortcut from the first layer to the last layer for each block. Not respecting blocks also significantly harms the final performance.

4.3 Models at Work

The eleven selected models are all implemented and the results are visualized. Due to limited space, let us take ConvNeXtBase as an example to show the implementation and result's visualization. The accuracy and loss of training and validation for ConvNeXtBase are shown in Figure 3. The confusion matrix is shown in Figure 4.

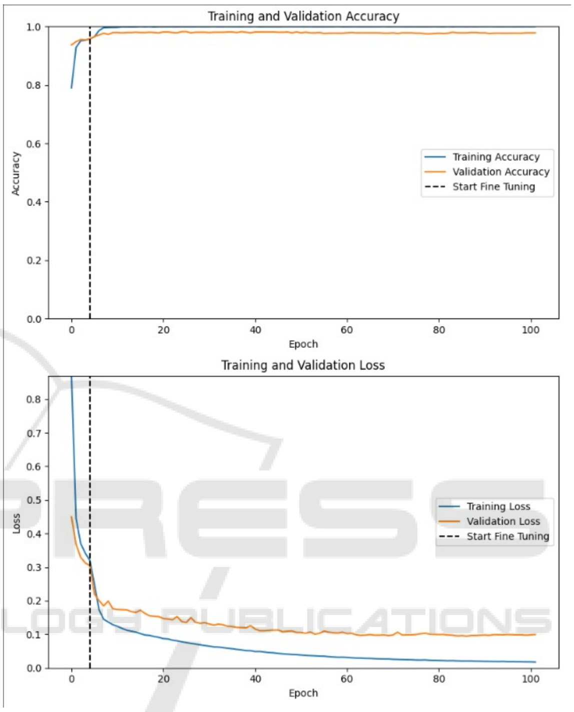

Figure 3: Accuracy and loss of training and validation for ConvNeXtBase.

4.4 Evaluation and Comparison

The accuracy, precision, recall and F1-score comparison for models on the test dataset is shown in Tables 3 to 6 respectively.

After comparing the performance of all models, ConvNeXtBase is considered as the best one, and EfficientnetV2L is the second best one for the specific classification task.

4.4.1 Comparison with Recent Results by Transfer Learning

Table 7 lists the accuracies obtained from some transfer learning studies in the past three years and compare them with this paper. Figure 5 shows visual comparison of transfer learning performance in this paper.

The comparative analysis of test results highlights the impact of dataset size, number of cate-

Table 3: Accuracy of models on test dataset.

Model	Accuracy	Model	Accuracy
MobileNetV2	94.68%	ResNet50V2	95.26%
VGG16	93.60%	InceptionResNetV2	96.05%
VGG19	95.02%	NASNetLarge	97.35%
DenseNet169	97.65%	**EfficientNetV2L**	**98.64%**
DenseNet201	97.71%	**ConvNeXtBase**	**99.00%**
Xception	96.42%		

Table 4: Precision of models for all categories on test dataset.

	batteries	brown glass	card-board	green glass	metal	paper	plastic	white glass
MobileNetV2	0.98	0.99	0.98	0.96	0.86	0.95	0.92	0.94
VGG16	0.99	0.99	0.97	0.96	0.90	0.96	0.87	0.91
VGG19	0.96	1.00	0.97	0.98	0.86	0.96	0.95	0.93
DenseNet169	0.98	1.00	1.00	0.99	0.93	0.97	0.96	0.98
DenseNet201	1.00	1.00	0.99	0.99	0.93	0.97	0.94	0.96
Xception	0.98	0.96	0.99	0.95	0.93	0.96	0.95	0.96
ResNet50V2	0.98	0.92	0.99	0.95	0.90	0.94	0.93	0.93
Inception-ResNetV2	0.98	0.94	0.99	0.95	0.93	0.97	0.94	0.92
NasNetLarge	0.98	1.00	0.96	1.00	0.93	0.95	0.97	0.97
EfficientNetV2L	**0.99**	**0.99**	**0.99**	**0.99**	**0.97**	**0.98**	**0.96**	**0.97**
ConvNeXtBase	**1.00**	**1.00**	**0.99**	**0.99**	**0.97**	**0.98**	**0.99**	**0.97**

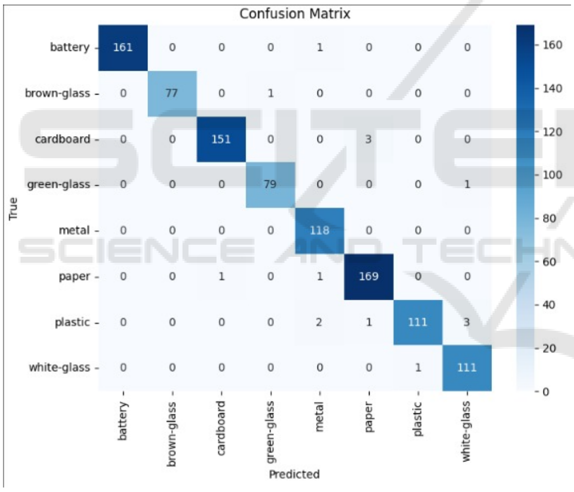

Figure 4: Confusion matrix for ConvNeXtBase.

gories, and model architecture on classification accuracy. ConvNeXtBase and EfficientNetV2L in this paper achieved 99.00% and 98.62% accuracy in 6531 images in 8 categories, respectively, showing excellent performance. This result demonstrates that the two models are able to efficiently capture complex features and have strong feature extraction capabilities. Also, the study reveals that while larger datasets generally enhance accuracy, the dataset quality and distribution are equally crucial. Models like VGG16 and MobileNetV2 demonstrated robustness across different datasets, but newer models such as EfficientNetV2L and ConvNeXtBase showed significant improvements, particularly with a moderate or large number of categories. The analysis underscores the importance of balancing dataset characteristics with the right model architecture to achieve optimal performance.

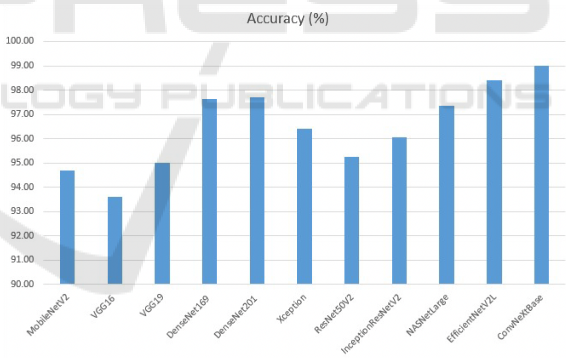

Figure 5: Accuracy comparison graphs for various transfer learning strategies.

5 CONCLUSIONS

This study evaluates several pre-trained CNN models and transfer learning techniques for classifying household waste into eight categories, analyzing eleven models with ConvNeXtBase and EfficientNetV2L as top performers, achieving 99.00% and 98.64% accuracy, respectively. The study's originality lies in the detailed comparison of models using expanded datasets and exploring fine-tuning techniques to enhance performance. Data augmentation

Table 5: Recall of models for all categories on test dataset.

	batteries	brown glass	card-board	green glass	metal	paper	plastic	white glass
MobileNetV2	0.98	0.94	0.95	0.99	0.97	0.98	0.91	0.86
VGG16	0.96	0.97	0.97	0.97	0.97	0.97	0.88	0.84
VGG19	0.96	0.94	0.96	1.00	0.97	0.96	0.88	0.92
DenseNet169	0.99	0.96	0.97	1.00	0.99	0.99	0.95	0.93
DenseNet201	0.99	0.99	0.97	0.99	0.99	0.99	0.95	0.92
Xception	0.99	0.95	0.95	0.97	0.97	0.99	0.92	0.93
ResNet50V2	0.99	0.90	0.95	0.95	0.96	0.98	0.86	0.91
Inception-ResNetV2	0.98	0.95	0.97	0.97	0.95	0.96	0.91	0.93
NasNetLarge	0.99	0.96	0.96	0.99	0.96	0.97	0.96	0.96
EfficientNetV2L	**0.99**	**0.99**	**0.99**	**1.00**	**0.96**	**0.99**	**0.97**	**0.93**
ConvNeXtBase	**0.99**	**0.99**	**0.98**	**0.99**	**1.00**	**0.99**	**0.95**	**0.99**

Table 6: F1-Score of models for all categories on test dataset.

	batteries	brown glass	card-board	green glass	metal	paper	plastic	white glass
MobileNetV2	0.98	0.96	0.96	0.98	0.91	0.97	0.91	0.90
VGG16	0.97	0.98	0.97	0.97	0.93	0.97	0.87	0.87
VGG19	0.96	0.97	0.97	0.99	0.91	0.96	0.92	0.92
DenseNet169	0.98	0.98	0.98	0.99	0.96	0.98	0.95	0.95
DenseNet201	0.99	0.99	0.98	0.99	0.96	0.98	0.94	0.94
Xception	0.98	0.95	0.97	0.96	0.95	0.97	0.94	0.95
ResNet50V2	0.98	0.91	0.97	0.95	0.93	0.96	0.89	0.92
Inception-ResNetV2	0.98	0.94	0.98	0.96	0.94	0.97	0.92	0.92
NasNetLarge	0.98	0.98	0.96	0.99	0.95	0.96	0.96	0.96
EfficientNetV2L	**0.99**	**0.99**	**0.99**	**0.99**	**0.97**	**0.98**	**0.97**	**0.95**
ConvNeXtBase	**1.00**	**0.99**	**0.99**	**0.99**	**0.98**	**0.98**	**0.97**	**0.98**

Table 7: Recent results achieved by transfer learning.

Authors, year	No. of images	No. of categories	Models	Accuracy
This paper, 2024	6,531	8	ConvNeXtBase	99.00%
This paper, 2024	6,531	8	EfficientNetV2L	98.62%
Haque et al., 2024	15,515	12	DenseNet169	99.58%
Lou and Gou, 2023	2,725	14	EfficientNetV2S	95.50%
Wu et al., 2022	7,000	4	VGG16	96.21%
Mehedi et al., 2023	24,705	2	VGG16	96.00%
Mehedi et al., 2023	24,705	2	MobileNetV2	95.51%
Shukurov, 2023	15,000	12	ResNeXt50	95.00%
Goel et al., 2023	21,984	10	EfficientNetB7	93.75%
Goel et al., 2023	21,984	10	DenseNet161	93.25%
Goel et al., 2023	21,984	10	InceptionV3	92.54%
Goel et al., 2023	21,984	10	ResNet50	94.70%
Ma et al., 2022	12,000	5	MobileNetV2	98.70%

and early stopping were used to avoid overfitting. The results underscore the superiority of modern neural network architectures in waste classification, emphasizing dataset quality and size. A comparative analysis with recent studies highlights the significance of dataset characteristics and confirms that larger, diverse datasets improve generalization and accuracy. Moreover, newer CNN models outperform older architectures in complex tasks. This study reveals the strengths and weaknesses of various models and underscores the importance of selecting the right pre-trained model, contributing to the current discussion on efficient waste classification.

REFERENCES

Chang, C. (2018). Garbage classification — kaggle.com. https://www.kaggle.com/datasets/asdasdasasdas/garbage-classification. Accessed 21-10-2024.

Endah, S. N., Shiddiq, I. N., et al. (2020). Xception architecture transfer learning for garbage classification. In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, pages 1–4. IEEE.

Géron, A. (2019). Deep computer vision using convolutional neural network. *Hands-On Machine Learning with Scikit-Learn, Keras & tensorFlow*, pages 445–496.

Goel, S., Markanday, S., and Gupta, A. (2023). A deep dive into accurate garbage classification using transfer learning techniques for environmental sustainability. In *2023 9th International Conference on Signal*

Processing and Communication (ICSC), pages 209–213. IEEE.

Haque, N., Toufiq, R., Islam, M. Z., and Shoukhin, M. A. A. T. (2024). Garbage classification using a transfer learning with parameter tuning. In *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, pages 1252–1256. IEEE.

Keras-Team (2024). Keras documentation: Keras applications — keras.io. https://keras.io/api/applications/. Accessed 21-10-2024.

Li, F.-F., Deng, J., Russakovsky, O., Berg, A., and Li, K. (2024). Imagenet — image-net.org. https://image-net.org/. Accessed 21-10-2024.

Lou, L. and Gou, N. (2023). Domestic waste image classification algorithm based on improved efficientnetv2. In *2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP)*, pages 1374–1377. IEEE.

Ma, H., Ye, Y., Dong, J., and Bo, Y. (2022). An intelligent garbage classification system using a lightweight network mobilenetv2. In *2022 7th International Conference on Signal and Image Processing (ICSIP)*, pages 531–535. IEEE.

Mehedi, M. H. K., Arafin, I., Hasan, M., Rahman, F., Tasin, R., and Rasel, A. A. (2023). A transfer learning approach for efficient classification of waste materials. In *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0636–0640. IEEE.

Melcher, K. and Silipo, R. (2020). *Codeless Deep Learning with KNIME: Build, train, and deploy various deep neural network architectures using KNIME Analytics Platform*. Packt Publishing Ltd.

Mohamed, M. (2021). Garbage classification (12 classes) — kaggle.com. https://www.kaggle.com/datasets/mostafaabla/garbage-classification. Accessed 21-10-2024.

Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

Ruder, S. (2017). Transfer learning - machine learning's next frontier — ruder.io. https://www.ruder.io/transfer-learning/. Accessed 21-10-2024.

Satvilkar, M. (2018). *Image based trash classification using machine learning algorithms for recyclability status*. PhD thesis, Dublin, National College of Ireland.

Shukurov, R. (2023). Garbage classification based on finetuned state-of-the-art models. In *2023 9th International Conference on Control, Decision and Information Technologies (CoDIT)*, pages 841–846. IEEE.

Wang, Z. (2022). Research on garbage image classification based on neural network. In *2022 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, pages 214–217. IEEE.

Wu, Y., Wang, Y., Wang, C., and Song, Y. (2022). A novel garbage images classification method based on improved vgg. In *2022 34th Chinese Control and Decision Conference (CCDC)*, pages 1571–1575. IEEE.