# WISE - Waste Image Sorting and Evaluation
## Ensemble CNN Pipeline for Trash Detection and Classification

Brett Castro
University of Virginia
gry6ks@virginia.edu

Tyler Gorecki
University of Virginia
ttg6nx@virginia.edu

Vishwanath Guruvayur
University of Virginia
vish@virginia.edu

Akaash Kamdar
University of Virginia
ak2znr@virginia.edu

Isaac Levy
University of Virginia
gbz6qn@virginia.edu

## Abstract

Effective waste management is a critical global challenge with far-reaching environmental, economic, and public health implications. Despite increasing awareness, less than 20% of global waste is currently recycled, resulting in significant resource depletion, environmental degradation, and public exposure to hazardous materials.

The **WISE** (Waste Identification and Sorting using Enhanced vision) project addresses this gap by employing deep learning for automated waste classification through image recognition. By enabling accurate and efficient real-time sorting, our system aims to optimize recycling processes, reduce human health risks, and facilitate the recovery of valuable materials. We conducted a comparative analysis of three state-of-the-art convolutional neural networks: ResNet, GoogLeNet, and EfficientNet—on a dataset of 19,762 labeled waste images spanning 10 distinct classes.

Our evaluation highlights each model's strengths and limitations in terms of accuracy, efficiency, and robustness, offering actionable insights for deploying AI-driven solutions in a wide range of real-world recycling infrastructures. This work demonstrates the transformative potential of AI in advancing sustainable waste management practices.

## 1 Introduction

Improper waste disposal leads to contaminated recycling streams, increased costs, and harm to ecosystems. Currently, annual waste generation exceeds 2 billion tons and is expected to increase by 70% to 3.4 billion tons by 2050. With this predicted drastic increase in global waste production, smart and scalable solutions will become even more essential.

Applications of automated waste sorting range from consumer waste sorting, e.g., municipal garbage and recycling collection, to large-scale and high stakes industrial waste management. In industrial applications, there is a need for more efficiently and accurately identifying not just waste or recycling, but also differentiating between distinct types of recyclable materials. "Conventional techniques like [the gold standard of] near-infrared spectroscopy (NIRS)...face difficulties in accurately classifying chemically similar samples, such as polyethylene terephthalate (PET) and PET-glycol (PET-G), which have similar chemical compositions but distinct physical characteristics" (3) and require different recycling processes. Previous work has shown, however, that these materials can be identified and separated

using deep learning computer vision.

This project leverages deep learning for automated waste sorting, aiming to improve the efficiency and accuracy of classification systems deployed in real-world scenarios. We will explain our processes and ideas for further progression as we strongly believe there is room for improvement using physical applications of these models in the field.

## 2    Related Works

Recent research by Chen et al. (2) demonstrated the use of ShuffleNet v2 for garbage classification tasks, showing that lightweight models can perform well on image-based waste categorization. Their work combines the ShuffleNet model and the depth-separable convolution method to create lightweight YOLOv5s for classifying waste items. Their tuned model, though, only contained 62% of the parameters from the original model. Our work builds on these foundations by evaluating a broader set of architectures and comparing them in terms of performance and applicability to real-world use cases.

## 3    Methodology

### 3.1    Dataset and Preprocessing

We used a labeled dataset consisting of 19,762 waste images, categorized into 10 distinct classes, including shoes, batteries, clothes, trash, glass, cardboard, metal, paper, plastic, and biological. We then manually classified these images into the following categories: *Recyclable*, *Organic*, *Specialty Recycling*, and *Non-Recyclable*.

The distinctions are shown below in Figure 1. Each image was manually annotated to ensure accurate ground truth for model training. Preprocessing steps included resizing images to a uniform shape, normalization, and data augmentation (random flips, rotations, and shifts) to enhance model generalization.



Figure 1: Item classification across garbage classes

### 3.2    Model Architectures

We evaluated three state-of-the-art Convolutional Neural Network (CNN) architectures, each representing a different, unique approach to balancing depth, efficiency, and representational power. These models were selected to explore architectures that have driven progress in image classification tasks:

- **ResNet-50**: ResNet-50 is a deep residual network known for its robust feature extraction using skip connections. With 50 layers in depth, this architecture contains residual blocks that address the vanishing gradient problem and enable the model to learn complex features from images.
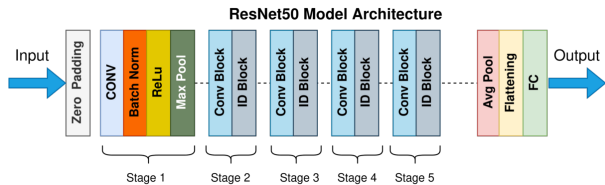


Figure 2: ResNet50 Architecture

- **GoogLeNet (Inception v1)**: Notable for its computational efficiency through inception modules, this architecture won 1st place at the 2014 ImageNet Large Scale Visual Recognition Competition. In

GoogLeNet, 1x1 convolution is used as a dimension reduction module to reduce computation, which can help increase depth and width.
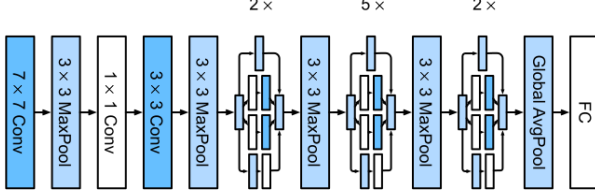


Figure 3: GoogLeNet Architecture

- **EfficientNet-B0**: This architecture is a compound-scaled model achieving a strong trade-off between performance and speed. Using compound scaling, this method is justified by the intuition that if the input image is bigger, then the network requires additional layers to increase the receptive field and requires more channels to caputer fine-grained details on the bigger image.
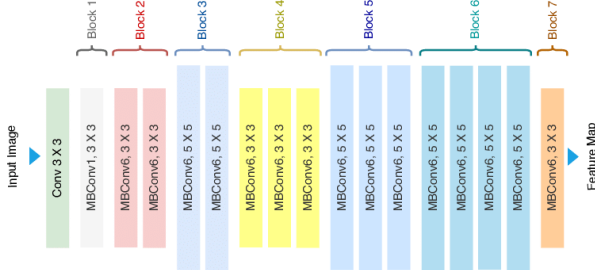


Figure 4: EfficientNet Architecture

Each model was trained using the same train-validation split and hyperparameter setup, ensuring a fair comparative evaluation. Even the Compute Power used were the same for each model's training phase.

## 3.3 Training Configuration

All models were trained on GPU-enabled hardware using the Adam optimizer and categorical cross-entropy loss. Early stopping and learning rate scheduling were applied to avoid overfitting and accelerate convergence.

# 4 Experiments and Results

The output from each of our models showed strong, encouraging results with high outcomes of all relevant metrics.

Table 1: Model Training Time

| Model | Training Time (s) |
| --- | --- |
| GoogLeNet | 760 |
| ResNet-50 | 2136 |
| EfficientNet-B0 | 2471 |

## 4.1 Evaluation Metrics and Insights

To assess model performance, we evaluated each CNN architecture using two key metrics: classification accuracy and training time.

- **GoogLeNet** was the *fastest model* to train, completing in 760 seconds (12.67 minutes), and showed the highest classification performance.

```
Classification Report (GoogLeNet):
             precision    recall  f1-score   support

     battery     0.936     0.992     0.964       133
  biological     0.978     0.978     0.978       137
   cardboard     0.989     0.945     0.966       289
     clothes     0.995     0.991     0.993       813
       glass     0.944     0.966     0.955       435
       metal     0.964     0.918     0.940       146
       paper     0.963     0.951     0.957       247
     plastic     0.931     0.931     0.931       303
       shoes     0.963     0.966     0.964       295
       trash     0.928     0.973     0.950       146

    accuracy                         0.966      2944
   macro avg     0.959     0.961     0.960      2944
weighted avg     0.966     0.966     0.966      2944
```

Figure 5: Detailed Classification Metrics for GoogLeNet across 10 waste categories.

- **EfficientNet** delivered *mediocre classification metrics*, and it also had the longest training time at 2471 seconds (41.18 minutes).

```
Classification Report:
              precision    recall  f1-score   support

     battery      0.869     0.974     0.919       116
  biological      0.915     0.952     0.933       147
   cardboard      0.915     0.881     0.898       295
     clothes      0.979     0.950     0.964       780
       glass      0.966     0.727     0.830       469
       metal      0.531     0.969     0.686       161
       paper      0.897     0.798     0.844       262
     plastic      0.747     0.845     0.793       297
       shoes      0.954     0.909     0.931       297
       trash      0.721     0.727     0.724       139

    accuracy                          0.871      2963
   macro avg      0.849     0.873     0.852      2963
weighted avg      0.894     0.871     0.876      2963
```

Figure 6: Detailed Classification Metrics for EfficientNet across 10 waste categories.

- **ResNet** achieved high accuracy, however took a long time to train at 2136 seconds (35.6 minutes).

```
Classification Report (ResNet50):
              precision    recall  f1-score   support

     battery      0.928     0.963     0.945       134
  biological      0.971     0.978     0.975       138
   cardboard      0.946     0.972     0.959       289
     clothes      0.989     0.987     0.988       818
       glass      0.944     0.957     0.950       439
       metal      0.902     0.884     0.893       146
       paper      0.955     0.943     0.949       247
     plastic      0.938     0.895     0.916       304
       shoes      0.942     0.970     0.956       299
       trash      0.951     0.906     0.928       149

    accuracy                          0.955      2963
   macro avg      0.947     0.945     0.946      2963
weighted avg      0.955     0.955     0.955      2963
```

Figure 7: Detailed Classification Metrics for ResNet50 across 10 waste categories.

These results help guide model selection based on use-case priorities—speed, accuracy, or a trade-off between the two. For real-world implementation, especially in environments with limited computational resources or real-time constraints, such insights are critical.

Table 2: Evaluation Metrics for CNN Models

| Metric | ResNet50 | GoogLeNet | EfficientNet |
|--------|----------|-----------|--------------|
| Accuracy | 0.955 | 0.966 | 0.871 |
| Precision | 0.947 | 0.959 | 0.849 |
| Recall | 0.945 | 0.961 | 0.873 |
| F1-Score | 0.946 | 0.960 | 0.892 |
| Support | 2963 | 2944 | 2963 |

## 4.2 Results Summary

The top-performing model in both evaluation metric categories, training time, and accuracy was GoogLeNet. ResNet slightly underperformed in accuracy and took a long time to train. EfficientNet performed the worst across the board.

These results indicate a trade-off between speed and accuracy in real-world deployment scenarios. Notably, inference time does not vary across models, indicating that with adequate available training and tuning resources, the computational overhead at time of use would likely be similar across the models.

## 5 Discussion

GoogLeNet emerged as the most balanced choice for deployment due to its superior classification metrics and quick training time. GoogLeNet's speed advantage may suit edge devices, where real-time analysis of object type detection and classification into garbage types is time sensitive. ResNet displayed high accuracy and may prove more useful with a different dataset. EfficientNet requires more tuning for improvement.

The performance of the models varied by class complexity, highlighting the need for robust architectures in applications such as smart bins or automated recycling plants.

Each of these models may do well in specific parts of the feature space of the images we have. So, having an educated ensemble of models would be beneficial to have an overall winner prediction framework.

# 6    Conclusion and Future Work

The WISE project validated the potential of CNNs for waste classification. ResNet's performance makes it suitable for real-world deployment, including smart waste bins and mobile applications.

Future work could explore further transfer learning, data augmentation, and real-time deployment using embedded systems.

For an industrial product out of this project, we will need to train the ensemble model on a high volume of images across various lighting settings to ensure that all edge cases are covered. Having confidence intervals of predictive capabilities will help users make an informed decision in case of ambiguous situations. Furthermore, industrial settings could benefit from the combined use of our deep learning image classifiers along with NIRS or other current gold standard methods to increase robustness and accuracy, particularly in high-volume and high-stakes environments.

A further extension would be to add live bounding box identification to correctly select which object in the image/video is the actual piece of waste. The model is also not currently set up to handle video frames, but in real-world settings, it would almost certainly be used with video, which adds further complexity and increases the necessity of more advanced segmentation and use of bounding boxes.

A model that is trained with real-world edge cases and real-time image identification methods as described above would allow us to build a real-time deployment system that could be displayed by public waste bins (airports, schools, malls, etc.). This application reduces the extensive need for sorting later on and speeds up the waste removal process. Along with improved accuracy, we would be able to prevent dangerous waste mishandling (such as throwing away batteries) and prevent recycling contamination that can occur if trash is wrongfully included. The same system could be shared as a mobile app, increasing consumer awareness while allowing better sorting to occur even without the advanced waste bin.

In addition, behind the scenes, recycling facilities can implement this technology to allow robotic arms to sort instead of the need for manual human sorting. This reduces the likelihood of human error and speeds up processes, again creating improved results and reducing risks to operators who might otherwise come into prolonged contact with hazardous materials.

Interesting considerations of applying these models in a production environment would be when to make an unsure prediction or ask the user for a better picture (angle, lighting, etc.). Asking for another picture is straightforward in the application mentioned above of a single public waste bin with a user interface, but in an industrial setting, this is not always possible and would decrease efficiency and usefulness. This would require further testing to see where the model inconsistencies lie.

If everything is below a certain confidence threshold, then in some cases nothing should be guessed, but other times it may be more nuanced and the model actually has inconsistent predictions on certain objects or types of images.

# References

[1] Aral, R. A., Keskin, S. R., Kaya, M., & Haciomeroglu, M. (2018). *Classification of trashnet dataset based on deep learning models.* 2018 IEEE International Conference on Big Data (Big Data)

[2] Chen, A., Liu, Y., & Zhang, Q. (2023). *Image-Based Waste Classification Using Lightweight CNNs.* International Journal of Environmental AI, 15(2), 122-135.

[3] Choi, J., Lim, B., & Yoo, Y. (2023). *Advancing plastic waste classification and recycling efficiency: Integrating image sensors and deep learning algorithms.* Applied Sciences, 13(18), 10224.