EcoSort: Garbage Classification

Hassan Qayyum
Student# 1008063836
hassan.qayyum@mail.utoronto.ca

Briana Vieira
Student# 1007790437
briana.vieira@mail.utoronto.ca

Areeba Mobeen
Student# 1007096786
areeba.mobeen@mail.utoronto.ca

Rayan Mustafa
Student# 1008224191
rayan.mustafa@mail.utoronto.ca

Abstract

The following provides an overview of the advancements made in the EcoSort project, aimed at developing an AI model for classifying garbage classes to facilitate efficient waste disposal. It highlights the progress achieved thus far, focusing on key aspects such as data processing and cleaning as well as current best models. We discuss the baseline model established to set a benchmark for the project's performance and elaborate on the current performance of the primary model, which leverages transfer learning to enhance classification accuracy.
—-Total Pages: 9

1 Introduction

The average person will spend one-third of their lifetime at work and at the same time generate 0.74kg of daily waste, contributing to roughly 2 billion tonnes globally every year (Li & Chen, 2023). Workplaces are challenged with managing a significant amount of waste, our team poses the question of how successfully this is being done. If executed correctly, waste sorting enables the recovery and recycling of valuable materials like metals, paper, and plastics, which conserves natural resources, reduces energy consumption, and lowers greenhouse gas emissions.

Currently, these systems are entirely manual and rely on the consumer to correctly sort their waste, and are consequently error-prone. As is demonstrated by the millions of tons of garbage that is not sorted correctly and end up in improper disposal sites, contributing to environmental degradation and climate change each year (Deer, 2021). In Toronto alone, the blue bin program manages 180,000 tonnes of recyclables annually, yet 30 percent (54,000 tonnes) is sent to landfills due to contamination (Zettler, 2019). This fully manual method has plenty of room for improvement and hinders progress toward efficient and accurate waste classification by not employing data-driven approaches.

Deep learning, particularly CNNs, offers a promising approach due to its ability to automatically learn discriminative features from raw data, making it well-suited for image-based tasks like garbage classification. Other machine learning approaches like RNNs, SVMs, and Random Forests are less suitable for image-based tasks such as garbage classification. RNNs are designed for sequential data tasks, while SVMs and Random Forests rely on handcrafted features or predefined rules, which may not capture the diverse visual characteristics of waste materials as effectively. Overall, CNNs offer a powerful and efficient solution for garbage classification, leveraging the capabilities of deep learning to automatically learn and extract discriminative features from raw image data.

Employing this CNN model approach, our project, EcoSort, addresses the pressing issue of improper waste management by developing a robust garbage classification system that will facilitate proper disposal efforts. Our model to classifies waste into the following six classes: cardboard, glass, metal, paper, plastic, and trash, of which the outputs can then be grouped into umbrella categories as needed. By leveraging the power of deep learning, we can enhance waste management practices, mitigate environmental pollution, and contribute to the fight against climate change.

2 ILLUSTRATION

For our final model, we employed transfer learning, utilizing the ResNet50 model to extract embeddings from the images. ResNet50, having been trained on extensive datasets like ImageNet, provides a robust feature extraction capability without the need to train new weights for feature extraction. Images get fed to ResNet50, then the extracted embeddings are subsequently input into our classifier, which then determines the class to which each item belongs. The high-level design of the EcoSort Model is illustrated in Figure 1, with a more in-depth discussion of the architecture provided in the following sections.

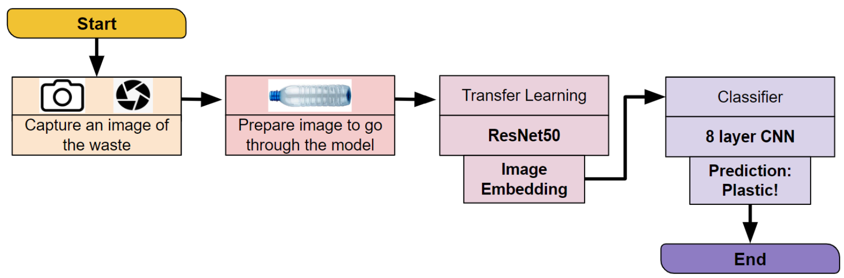


Figure 1: Design and flow of EcoSort in action

3 BACKGROUND AND RELATED WORK

Automation is becoming the norm in many aspects of life, yet it is not used in systems for the disposal of our daily waste. Let's take the manual method used at the University of Toronto as an example, in 2022 the St. George campus only captured 75% of recyclable material, and of the material placed in recycling, 4% did not belong there (Schwalb, 2023). Additionally, a study done at Dalhousie University showed that 11.5% of people do not sort their waste correctly (Smith et al., 2018). Clearly, there is room for improvement in this manual system for waste management, and the City of Toronto estimated that each percentage point decrease in contamination could lower recycling costs in the city by $600,000 to $1 million a year (Jonas, 2019).

Other groups have worked to develop deep learning models which can classify different types or components of waste. CNN models were found to be the most effective models for classification of waste by a thesis student from Tampere University (Bhandari, 2020). This thesis also discussed models that binarily sort images, one which classifies if the material in an image is trash or not and another which classifies if it is recyclable. There have also been models which classify images into multiple classes, for example 3 (disposal, recycling, and paper) and another which used 6 (cardboard, metal, glass, plastic, paper, and organic waste). The model with 6 classes achieved 95.46% accuracy and used a combination of CNN and graph long short-term memory (Li & Chen, 2023). There are also some approaches which use pre-trained ResNets, the best of which differentiated between waste that is either biodegradable or not and achieved 98.01% accuracy (Md.Rahman et al., 2022).

The current waste management system at U of T of four categories (landfill garbage, paper, containers, and coffee cups) lacks information to confirm whether it is effective. Additionally, our team could not find a model that uses the signs at the garbage bins to determine where users should place their waste. With the information and research we currently have, we aim to develop a successful model for garbage classification.

4 DATA PROCESSING

To collect data, our team used three large datasets from Kaggle. These combined datasets provided us with over 23,000 images, enough to ensure a robust model.

4.1 Preliminary Data Cleaning

We consolidated three datasets, each containing more classes than required, into a single dataset with six folders, each representing a class. To achieve this, we developed a Python script to categorize the files into their respective class folders, discarding any unnecessary files in the process. After consolidating all images into a single folder, creating an 'ImageFolder' from that directory was straightforward. We decided to preview some of the images by viewing them to see the quality and although we did not go through all 23,000 images, we quickly realized that there were many duplicates and some images that would not be useful as they were very dark or outliers.

4.2 Duplicates and Outliers

To ensure the integrity of our model, we took measures to minimize duplicates and outliers in the final dataset. Outliers were images that did not fit any garbage class or were excessively blurry, dark, or unclear. To assist us in this task, we leveraged the use of the FastDup tool because it provided a fast and reliable way to detect duplicate images within our dataset. By doing so, we were able to identify and remove outliers and duplicate images, ensuring that our model trained on a clean and representative dataset. This tool helped us identify almost 13,000 images that were not of adequate quality. Figure 2 shows the data distribution before and after cleaning. Overall, the use of FastDup was instrumental in ensuring the integrity of our dataset and the robustness of our model as it helped us polish our data.

Class	Before Cleaning		After Cleaning	
Cardboard	4038	16.89%	1606	14.49%
Glass	4067	17.01%	2682	24.19%
Paper	3458	14.46%	1025	9.25%
Plastic	4369	18.27%	1746	15.75%
Metal	4371	18.28%	1730	15.61%
Trash	3610	15.10%	2296	20.71%
Total	**23913**	**100.00%**	**11085**	**100.00%**

Figure 2: Data distribution before and after cleaning

4.3 Data Augmentation and Balancing

In Figure 2, it's evident that despite cleaning the data, a significant imbalance persists, with the metal class being notably underrepresented compared to the dominating glass class. To mitigate this, we employed data augmentation techniques to further augment the size of these classes.

Data augmentation serves a dual purpose, it not only helps in creating a more balanced dataset but also enhances the model's robustness. By exposing the model to a variety of augmented images, including challenging ones, the model becomes better equipped to learn and generalize effectively.

We applied a standard set of augmentations to all classes, including random rotations, horizontal flips, color jitter, resized crops, and Gaussian blur. This diverse range of transformations significantly increased the model's exposure to varied data, enhancing its ability to handle different scenarios. Furthermore, this process also effectively doubled the size of the dataset.

Despite the overall augmentation, the metal class remained the most underrepresented. To address this, we applied an additional set of augmentations specifically tailored to the metal class. These included random rotations between -10 and 10 degrees, color jitter, random affine transformations, and flips.Figure 3 illustrates the data distribution after applying these augmentations.

To balance the dataset, we implemented a random image removal strategy across all classes to match the number of images in the least represented class, which in this case was the metal class. While this approach resulted in the loss of some data, our dataset was large enough to maintain model robustness. Importantly, we ensured that the removal process was random to avoid disproportionately affecting certain classes, such as removing a large number of augmented images. Figure 3 illustrates the final distribution of the data after this balancing step.

Class	After Augmentation		After Balancing	
Cardboard	3212	14.49%	3075	16.67%
Glass	5364	24.19%	3075	16.67%
Metal	2050	9.25%	3075	16.67%
Paper	3492	15.75%	3074	16.66%
Plastic	3460	15.61%	3075	16.67%
Trash	4592	20.71%	3075	16.67%
Total	**22170**	**100.00%**	**18449**	**100.00%**

Figure 3: Final dataset distribution

4.4 Splitting the Data

To split the dataset into training, validation, and testing sets, we utilized the `train_test_split` function from the `sklearn` library as it preserves the distribution of classes across each dataset, ensuring the model training is as unbiased as possible. We chose a split ratio of 70%, 15%, and 15% for training, validation, and testing datasets, respectively. This equates to 12,194 number of images in training, 2767 in validation, and 2768 in testing.

5 Architecture

EcoSort's task is to classify images of garbage into six classes: cardboard, glass, metal, paper, plastic and trash. The type of model most well suited for image classification is a Convolutional Neural Network (CNN) as it is able to learn distinct and recognizable features wherever they may appear on the image. This is done by utilizing mathematical constructs such as convolution filters and pooling.

In the past, others have made successful CNNs which are very talented at extracting features. One of the first of these models that was developed is AlexNet, which has over 62.8 million training parameters [ref A]. A more recent model is ResNet which is able to go much deeper and combat vanishing gradients by leveraging Residual Networks and incorporating skipping connections. To utilize this, we employ transfer learning. Transfer learning is a more efficient approach in this context because it leverages the features learned by these pre-trained models. These models are trained on large datasets like ImageNet, which enables them to capture generic features such as edges, textures, and shapes that are useful for a wide range of computer vision tasks. By using transfer learning, we can start with these pre-trained models and fine-tune them on our specific dataset, allowing us to achieve good performance with less computational resources and training time compared to training a model from scratch. Additionally, transfer learning helps prevent overfitting, as the pre-trained models have already learned meaningful representations from vast amounts of data.

After experimenting with several configurations, we settled on a classifier comprising three convolutional layers that utilize 3x3 filters. This choice allows us to emulate the effects of larger resolution filters while minimizing the number of adjustable parameters. As another result of hyperparameter search, we chose six fully connected layers follow the three convolution layers; a higher number of layers significantly increased the number of trainable weights and fewer layers decreased the model's performance. In addition, We adopted the bottleneck and max pooling approach to enhance the model's feature extraction and selection process. Max pooling and bottleneck techniques expand and then shrink the feature maps which compel the model to retain only the most critical features.

An overview of the architecture and flow is depicted in Figure **??**. We leveraged the 49 convolution layers (discarding the final fully connected layer) and approximately 25 million parameters from ResNet50, importantly, we do not have to train these layers, thanks to the advantages of transfer learning (Keras 3 API documentation). The classifier itself consists of 7,078,912 trainable weights from the convolutional layers and 1,287,270 weights from the five fully connected layers, totaling about 8.366 million trainable weights. The model was trained using various hyperparameters, and the optimal configuration was found to be a learning rate of 0.0001, a batch size of 128, 50 epochs, and the Adam optimizer.

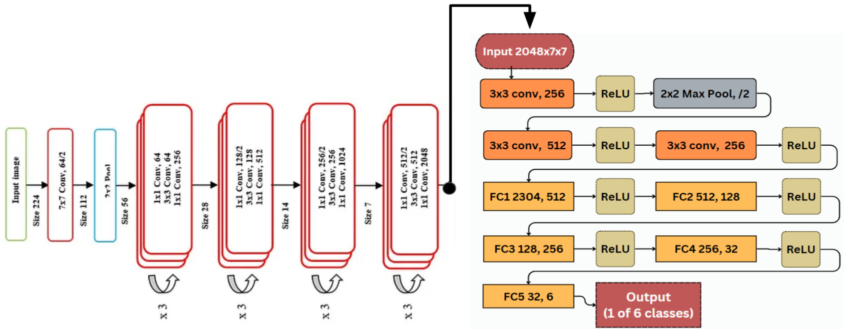


Figure 4: EcoSort model architecture

6 Baseline Model

EcoSort aims to classify images into six classes of garbage: cardboard, glass, metal, paper, plastic, and trash. Therefore, a Convolutional Neural Network (CNN) is the most suitable baseline model. CNNs excel at feature learning in classification tasks by employing mathematical constructs such as convolution filters and pooling. These techniques enable the model to discern distinct and recognizable features within smaller sections of the data.

Our team chose a 5-layer CNN as a baseline model for our image classification project. The architecture of the model is depicted in Figure 5. We chose this simple CNN as a starting point to establish a performance benchmark to which we can compare more complex models' performances. Our primary expectation from the baseline model is that it can generalize and accurately classify images in our datasets.

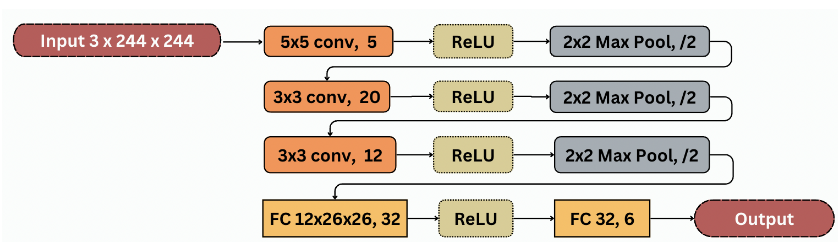


Figure 5: Baseline model

We used three convolutional layers, each layer further capturing more complex patterns to produce feature maps of the input images. Our chosen kernel size is 3x3 since any large kernel can be approximated with 3x3 kernels, ultimately being the most efficient choice as shown by GoogLeNet. The first kernel is 5x5 to resize the input image as it is convenient to use a kernel of this size on the first layer. Additionally, a max pooling layer is used in between each convolutional layer to reduce the spatial dimensions of the feature maps. Finally, the convolution and pooling layers are followed by two fully connected layers for our model's purpose of image classification.

6.1 Comparing the Baseline Model with the Primary Model

To evaluate the accuracy of our primary model, we will compare it to the benchmarked baseline model. The baseline model achieved a validation and test accuracy of 53% and a test accuracy of 50%. Only testing the performance accuracy of the primary model is not robust enough, as we cannot conclude if high accuracy results are from well-performing classification or the simplicity of the problem. The CNN as a baseline model offers us a straightforward and computationally efficient comparison point and requires minimal fine-tuning. By comparing the primary model to our benchmark model, we can evaluate it based on how well it outperforms the baseline model. If it is not much better than the baseline model, it suggests that it is not performing well and not meeting our expectations.

7 Quantitative Results

Let us now analyze the effectiveness of the EcoSort garbage classification model by examining quantitative metrics such as accuracy, loss, precision, and recall. The final model was able to achieve a training accuracy of 99% and a validation accuracy of 82%, as shown in Figure 6. However, by looking at the accuracy and loss curves in Figure 6 we can see that while the training accuracy keeps increasing as the number of epochs increases, the validation accuracy seems to plateau. This aligns with how the training loss continues to decrease while the validation loss starts to increase as well at this point.

Figure 6: Accuracy and Loss training curves over fifty epochs

The discrepancy between the training and validation datasets indicates the model is overfitting. This means that the model is learning to memorize the training data, rather than learn patterns and be able to generalize to unseen data. However, since the validation accuracy is just staying the same and not decreasing, it appears that the model is still becoming better at classifying some patterns.

A more granular analysis of the model's performance can be obtained by investigating the confusion matrix of the validation dataset. From this, our team was able to calculate the prediction recall and precision of each class which is shown in Figure 7.

Class	Recall	Precision
Cardboard	83.09%	84.62%
Glass	86.48%	84.42%
Metal	89.57%	83.40%
Paper	84.57%	82.82%
Plastic	71.49%	83.13%
Trash	84.29%	81.14%

Figure 7: Recall and precision values for each class

For the cardboard and plastic classes, the model performed better on precision than recall, and the opposite is true for glass, metal, paper, and trash. High precision and low recall implies that there is low chance for the model to predict an image belongs to any of these classes but high chance that if it predicts these classes it will be a true positive. Low precision and high recall implies the inverse.

8 Qualitative Results

After the training phase of our final model, we evaluated how well the model was able to perform on images in the validation data set of different classes. From the model's confusion matrix in Figure 8, we observed how the model consistently performed well across all 6 classes. A value closer to 400, signifies a higher number of correct predictions made by the final model. As such, the highest number of correct predictions for each true label row is on the cell with its corresponding predicted class, indicating accurate predictions for the class. Additionally, we noted that confusion arises between plastic and trash, plastic and glass, as well as cardboard and paper. The pair of classes our model confuses the most is plastic and trash, with a cumulative of 60 images being classified as the opposite class. Our analysis of why this occurs is discussed in Section 10.

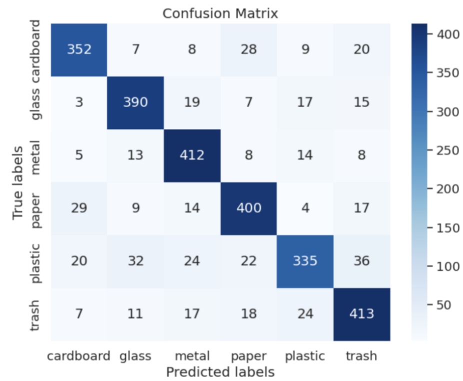


Figure 8: Confusion matrix on validation dataset

We displayed several images of various classes that our final model classified, along with their ground truth labels. The results are shown in Figure 9, and by visually comparing the truth label to the predicted class, we can observe whether or not the final model performed a correct classification. We concluded that our model was able to classify well, as 7 out of the 9 images displayed were classified correctly. Additionally, we noted that confusion arises between cardboard and paper here as well, as observed in our confusion matrix results. It is also important to note that this is not the most reliable measure of accuracy as this is only a small output of 9 images, but it gives a good visual representation of how well the model performs on this small subset.

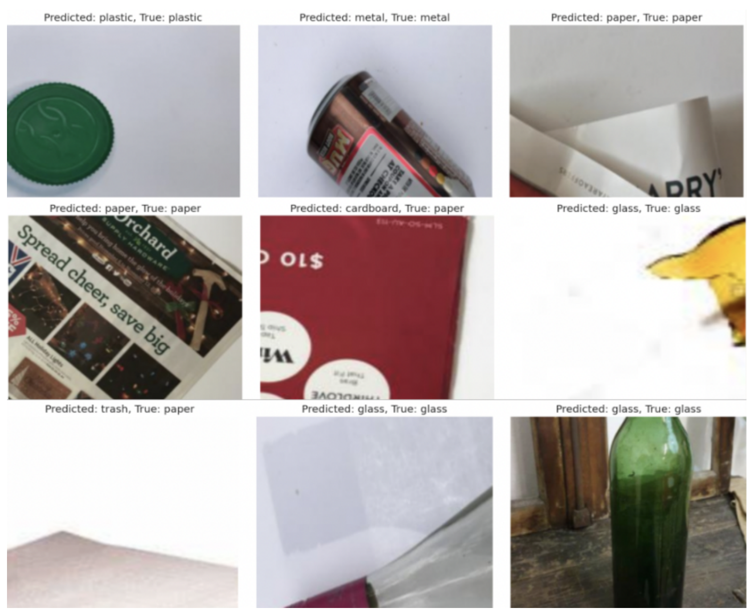


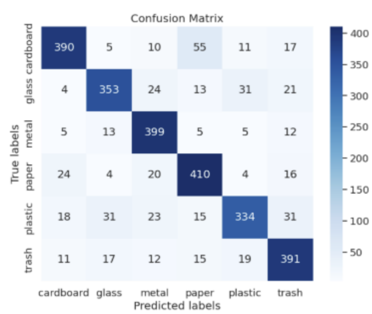
Figure 9: Model's classifications with truth labels

9 Evaluate Model on new Data

During data processing, our team set aside 2800 images as a test dataset, which were not to be used during the development of our model. By taking this precaution, our team can ensure that the test results we present are a good representation of the model's performance on new data and are an indicator of how EcoSort would perform in practice.

EcoSort's final model achieved an 82% accuracy on this test dataset, with performance comparable to the validation set. The confusion matrix in Figure 10 highlights the model's ability to generalize well to unseen images across different classes. The model's highest recall was for the metal class, indicating its effectiveness in correctly identifying metal items. Conversely, the highest precision was for the cardboard class, suggesting that cardboard predictions were more likely to be correct. However, the model struggled with the plastic class, showing the lowest recall, and the paper class, with the lowest precision. Similar to the validation dataset, the model continues to confuse paper and cardboard classes, as well as glass and plastic. Our analysis of why this occurs is discussed in Section 10. Despite these challenges, the overall performance of the model on the test dataset demonstrates its robustness and ability to generalize effectively.

To further demonstrate the proficiency of the model, we tested the model prediction against a single image at a time that we captured. For example, we input an image of a plastic water bottle and print the model's prediction. The model correctly classified it as plastic. We tried this several times with different images and the model continued to predict correctly, highlighting the model's ability to perform on raw data.



Class	Recall	Precision
Cardboard	79.92%	86.28%
Glass	79.15%	83.45%
Metal	90.89%	81.76%
Paper	85.77%	79.92%
Plastic	73.89%	82.67%
Trash	84.09%	80.12%

Figure 10: Confusion matrix analysis on test dataset

10 Discussion

Not only did we want to be better than our baseline model, we aimed to have a higher accuracy than the current system in the real world as well. The current waste management system relies on humans to make classifications, and at the University of Toronto, it was found that only 75% of recycling material is being captured correctly. Our model achieving a test accuracy of 82% exceeds both of these expectations.

As discussed in the previous sections, confusion arises between the following classes; plastic and trash, plastic and glass, as well as paper and cardboard.

Plastic being mistaken for trash is bad because plastic items should ideally be recycled, though it is worse that our model is mistaking trash for plastic. This results in contamination, which would cause all of the recycling to be discarded as trash, defeating the purpose. After examining the misclassified images, we discovered that the model often misclassified crumpled water bottles as trash, likely because their plastic nature was difficult to discern. Additionally, instances where trash was misclassified as plastic often involved colorful toothbrushes standing upright, which could be mistaken for plastic bottles due to similarities in shape, color, and orientation. These findings parallel human intuition, as individuals are more likely to discard a water bottle in the trash than in the metal or glass bins, for example.

Plastic and glass have very different recycling processes, thus confusing them for each other resulting in contamination on both sides. This is a pitfall of our model, as humans can easily discern between these materials and are not likely to make this mistake when discarding waste. After examining images where glass items were classified as plastic, we found that the model often confused tall glass containers with tall plastic bottles, leading to misclassification. The similarity in appearance between these objects likely contributed to the confusion, causing the model to classify them as the wrong class. This challenge highlights the inherent difficulty for deep learning models to understand the material composition of objects, as their learning primarily relies on distinct physical features rather than material properties.

Upon reviewing images where paper items were misclassified as cardboard, we found that the model struggled to differentiate between the two materials, likely due to their similar appearance and texture, especially in flattened or folded forms. Cardboard items were sometimes misclassified as paper, particularly when they had paper labels on top or when the image showed only flat cardboard on a white background. However, in Toronto all types of paper, including cardboard, get recycled together as part of the blue bin program (Acceptable Recyclable items). Therefore, it is not an issue that our model confuses these classes.

These findings emphasize the need to enhance the model's capacity to differentiate between subtle visual cues, particularly in challenging scenarios where items exhibit diverse appearances. This is inherently difficult for deep learning models as their learning primarily depends on distinct physical features rather than material properties.

Figure 11: Plastic mistaken as trash, glass mistaken as plastic, cardboard mistaken as paper

11 Ethical Considerations

EcoSort raises ethical concerns, including privacy issues and geographical bias in the training data due to varying waste materials worldwide. Images used for training may contain personally identifiable information, such as addresses on recyclable letters. Consent is crucial, especially if faces are captured during the disposal process. These issues must be addressed to ensure responsible data handling. Additionally, our training data is limited geographically, leading to biases in the model. Different regions have unique waste practices and regulations, affecting the model's accuracy. We've attempted to mitigate this by balancing our dataset and removing duplicates. However, our model is optimized for municipal waste in Toronto and may struggle in unfamiliar regions or scenarios, potentially leading to misclassification and improper waste management, further exacerbating environmental issues.

12 Project Difficulty

As discussed in the report, we encountered several challenges that hindered our model's performance and increased difficulty in achieving high accuracy. These include difficulties in data cleaning and balancing, confusion among classes, and limited computational resources. The computational resources required to extract the embeddings include needing over 30 GB of RAM, almost three times the amount available in Google Colab. Despite all the challenges, we exceeded our team's expectations for this project and developed a model which decreases the cost, increases efficiency, making it easy for the user to sort their garbage correctly.

13 GitHub

The following repository (link to repository) contains all the Google Colab notebooks used to make progress, including those for data processing, the baseline model, and the final model files.

REFERENCES

Acceptable Recyclable items. *What goes in the blue bin prohibited items*.

Sishir Bhandari. Automatic waste sorting in industrial environments via machine learning approaches. *Tampere University*, 2020.

Ryan Deer. Landfills: We're running out of space. *ROADRUNNER*, 2021.

Sabrina Jonas. Garbage or recycling? u of t students invent 'robobin' to make the decision for you. *Canadian Broadcasting Corporation*, 2019.

Keras 3 API documentation. *Keras Applications*.

Ninghui Li and Yuan Chen. Municipal solid waste classification and real-time detection using deep learning methods. *Urban Climate*, 2023.

Wahidur Md.Rahman, Rahabul Islam, Arafat Hasan, Nasima Bithi Islam, Mahmodul Md.Hasan, and Mohammad Motiur Rahman. Intelligent waste management system using deep learning with iot. *Elsevier B.V.*, 2022.

Jessie Schwalb. The breakdown: Recycling at u of t. *The Varsity*, 2023.

Laura Smith, Jessica Tam, Phil Martel, and Jeremy Holleran. Waste management at dalhousie university, canada: Social factors affecting recycling behavior and use of the four bin system at the studley campus. 2018.

Melanie Zettler. Toronto recycling: Why so much material still goes to landfill. *Global News: Environment*, 2019.