Recycle or not? Classifying Waste Items

Group 1: **Varun Pavuloori**, Erin Moulton, Hank Dickerson

Project Details

Context

- In the United States, the recycling rate is around 32%, increase from 7% in 1960 [1]
- Approximately 80% of what Americans throw away is recyclable
- As waste production grows with population and consumption, better solutions are needed to handle the strain on recycling infrastructure

Motivation

- The discrepancy between the US recycling rate and the amount of recyclable waste that is discarded necessitates a simpler method to determine trash vs recycling
- Improper sorting of waste contaminates recycling streams, increases operational costs, and reduces the efficiency of facilities, undermining the effectiveness of recycling initiatives.

Project Details

Hypothesis

A convolutional neural network (CNN) trained on a dataset of images of waste items will be able to accurately classify items into recyclable and non recyclable waste based on the visual features of the items.

Research Question

Can a machine learning model accurately classify images of waste into recyclable and non-recyclable categories based on visual features?

Modeling Approach

- Split image data into training, validation, and test sets
- Pretrained convolutional neural network (CNN model) to predict the recyclability of waste items [2]
- Validation accuracy to determine the success of the models predictions

Data

- **Acquisition**: Our dataset was sourced from a Github repository (sam-single) [3]
  - Each image sample captured from a Australian landfill site
- **4,752** usable images:
  - Images include cardboard, food organics, glass, metal, paper, plastic, textiles, vegetation, and miscellaneous trash.
- **Data Establishment:**
  - Cloned the repository, categorized images as recyclable/non-recyclable (label column), and classified them by waste type (waste_type column)

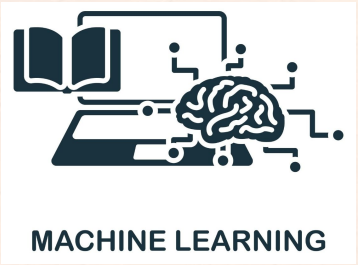
data dictionary

Column	Description	Type
image	524 x 524 resolution waste image samples	object
label	Labels the image as either recyclable or non recyclable.	int64
waste_type	Label of the type of waste	object

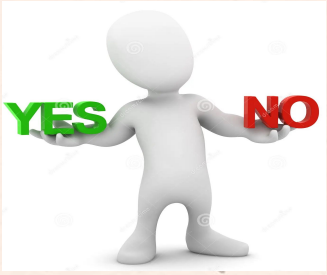
Analysis Plan & Justification

-Resized images to 224x224 pixels
-Normalized pixel values to a range of 0-1
-Data augmentation: rotation, flipping, & scaling

Label the images as either recyclable/non recyclable based on waste type



-convolutional neural network trained on the labeled dataset
-split into 70% training, 15% validation, and 15% testing



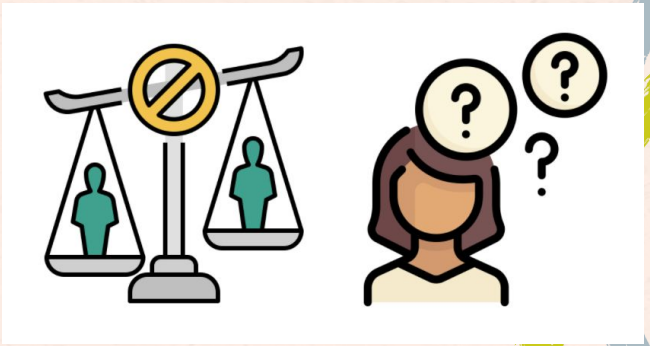
If the CNN achieves an accuracy ≥85% in classifying recyclable and non-recyclable waste, we will fail to reject our hypothesis

Tricky Decisions

- We faced a choice between designing our model to classify images as compostable, recyclable, and non-recyclable
  - We decided to focus only on recyclable vs non-recyclable due to the nuances of classifying items as compostable
    - For example, colored and glossy paper is not compostable which would have been difficult to train design our model to detect
  - Also, from our dataset, only food organics and vegetation are typically compostable items, making it challenging to train a reliable model on such few images

Bias & Uncertainty Validation

- Varying number of images per category
  - Approximately twice as many images in the recyclable subfolder compared to the non-recyclable subfolder
  - Certain materials, such as plastic and metal, had significantly more entries than others, further contributing to the uneven distribution
- Potential for Misclassified Images
  - Due to time constraints, we were unable to manually review all the images to confirm their correct categorization
  - As a result, some images may have been misclassified, which could potentially impact the accuracy of our results

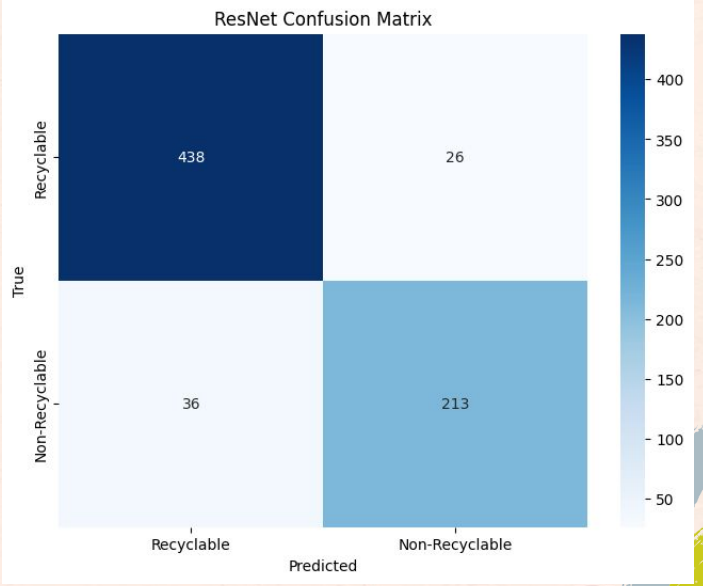


Results & Conclusion

Homemade CNN Model: 0.73 Accuracy; 0.82 Val Accuracy
loss: 1.1995; val_loss: 1.1242

Pretrained ResNet Model: 0.91 Accuracy; 0.91 Val Accuracy
loss: 0.0748; val_loss: 0.0749

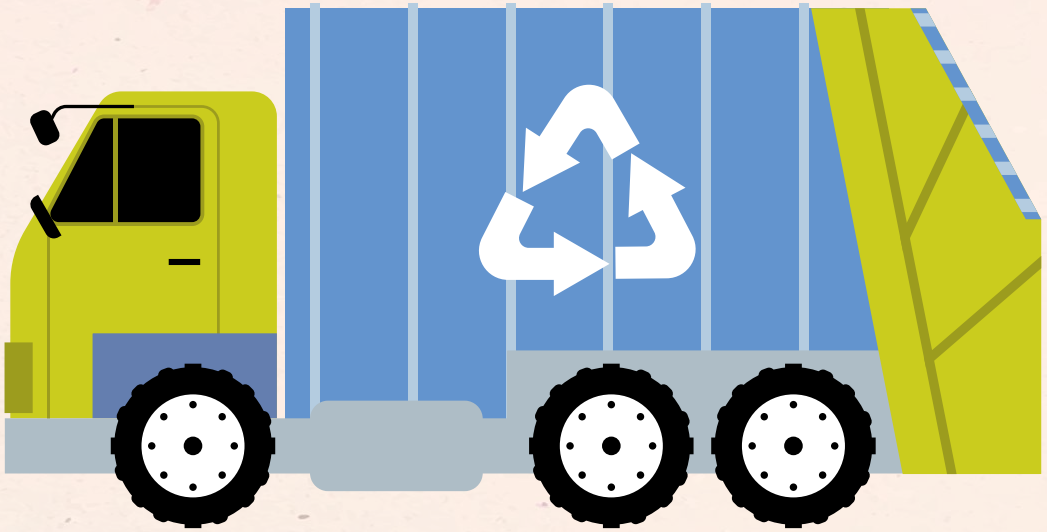
Reject the Null Hypothesis as ResNet is more than 85% accurate



Next Steps

- Manually evaluate our model predictions with the image it is classifying
  - Explore the decisions made and ensure that they align with the recycling standards in the US
- Test model classification on our own images to ensure that the dataset did not contain any hidden sources of bias (lens, background, etc)
- Go back and include functionality for compostable recognition
- Explore data to delve into subgroups of recyclable vs non recyclable plastic
- Potentially connect model with a camera for continuous live detection

Questions?



References

[1] U.S. Environmental Protection Agency. (n.d.). *America Recycles Day*. EPA. https://www.epa.gov/circulareconomy/america-recycles-day

[2] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). "ImageNet: A Large-Scale Hierarchical Image Database." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[3] sam-single. (2023). *GitHub - sam-single/realwaste: RealWaste is an image dataset of waste items in their authentic state.* GitHub. https://github.com/sam-single/realwaste?tab=readme-ov-file