Garbage Classification System Design Proposal

Course Name: Data Mining and Machine Learning
Professor's Name: Dr. Roberto Souza

Group <7> Members

	Name	ID
1	Mohammad Alhashem	30272178
2	Yuhao Huang	30284086
3	Noureldin Amer	30119675
4	Ali Karimi	30225064

30th of January 2025

1 Introduction

Garbage collection, classification and handling have been always a concern for many cities. In this document, a complete system to classify garbage for the city of Calgary based on cellphone photos and natural language descriptions will be proposed. The system aims to ensure specific sorting into the city's "Green", "Blue", "Black" bins, or categorize items as "Other" where applicable. Effective garbage classification promotes recycling, reduces landfill waste, and supports sustainability goals. This report outlines the design specifications, dataset creation strategy, and challenges associated with the system.

2 Problem Definition

This project aims to develop a garbage classification system that takes an image and a short text description as input to predict the correct bin for disposal. The task is a multiclass classification problem that integrates both images and text. Therefore, our experimental design follows a multimodal classification approach.

3 Methodology

3.1 Data Collection

The raw data we collect for training the system consists of two parts as follows:

- **Cellphone photos:**
  1. Each photo contains only one item to help the model focus on classification and reduce preprocessing steps such as segmentation.
  2. Objects are centered in the photos to aid feature extraction.
  3. Both homogenous and nonhomogeneous photos will be included in the dataset in order to increase the robustness of the system and ensure practicality and usability.
  4. We expect the dataset to include more than 10,000 images to ensure sufficient training data while maintaining diversity for better model generalization and robustness.
- **Natural Language Descriptions**:
  1. Each photo has an accurate and straightforward (3-8 words) description.
  2. Example descriptions: "dirty_foil_food_tray" or "used_tissue".

Photos and descriptions will be balanced across four categories: "Green", "Blue", "Black", and "Other" bins, as per Calgary's guidelines [1]. Balancing ensures that no category dominates the dataset, which could lead to biased predictions. Collecting such a scalable image data set can be achieved by different approaches such as automated camera systems at waste facilities and encouraging residents to upload labeled images on specific platforms.

3.2 Data Preprocessing

Different steps will be followed to enhance the data quality:

1. **Image Resizing**: All the images will be resized to 224*224, which is the expected input size of MobileNetV3 [2]. Specific and standard dimensions for the model input will ensure compatibility with possible neural networks architectures.
2. **Augmentation**: Data augmentation uses transformations such as rotation, translation, flipping, and scaling to enhance dataset diversity and improve model generalization.

3. **Normalization**: For a better convergence and to standardize the data, pixel values will be scaled to 0-1 range while unscaled for show purposes.
4. **Image Brightening:** by using automatic threshold-based method [3] we can adjust brightness for consistent lighting it also helps eliminate unwanted noise caused by lightning variations.

5. **Canny Edge Detection:** removes blank backgrounds, focusing on the main object in the image [3].
6. **Text:** We will first convert all text to lowercase**,** then correct any misspellings, and finally remove special characters, extra spaces, and non-alphanumeric symbols. We do not perform manual tokenization, stop word removal or stemming/lemmatization, as MobileBERT [4] inherently handles these processes during training.

We acknowledge that some of these steps are computationally costly, so we will start from a simple model with simple preprocessing. Then, if we need more accuracy, we will add the ignored steps (especially steps 3 and 4)

3.3 Model Selection

For this classification task involving both images and text input data, we adopt a multimodal learning architecture. Our proposed model consists of two parallel feature extraction modules processing images and text data separately, followed by a Dense Neural Network (DNN) which predicts the garbage category from the fused feature vector.

Considering that the model will be deployed on mobile devices, we prioritize both accuracy and inference time when selecting the base architectures. For image feature extraction, we use pretrained MobileNetV3 [2], as it is proven to offer high accuracy for garbage classification tasks while maintaining low latency [5]. For text feature extraction, we choose MobileBERT [4], which delivers performance comparable to BERT while being optimized for mobile deployment.

3.4 Model Evaluation

To assess the multimodal garbage classification model, we prioritize metrics that address class imbalance and ensure fairness across underrepresented categories. Balanced Accuracy (average per-class recall) mitigates bias toward dominant classes, while Macro F1-Score—a harmonic means of class-wise precision and recall—ensures equal weighting of all categories, critical for rare or ambiguous cases. Accuracy provides an intuitive overall correctness measure but is interpreted cautiously due to potential skew from majority classes. These metrics collectively evaluate robustness, with F1 inherently incorporating precision and recall, thereby capturing both false positives and false negatives. These metrics, summarized in the formula table below, align with best practices for multi-class evaluation [6].

4 Potential Challenges

1. **Device Variability**: Images will have different qualities since they will be taken from multi cellphone models which can impact consistency.
2. **Lighting and Backgrounds**: Ensuring consistent lighting will be difficult to maintain. Preprocessing steps like normalization can help address this matter.
3. **Class Overlap**: Items with features relevant to multiple bins such as greasy paper and clean paper. Specific train dataset will help the model to distinguish.
4. **Generalization**: Ensuring the model performs well on unseen data requires diverse and representative training data.

Metric	**Description (4-Class Garbage Context)**	**Formula (4-Class Garbage Context)**
Accuracy	Proportion of total correct predictions.	$\frac{\sum_{k=1}^{4} TP_k}{\text{Total Samples}}$
Balanced Accuracy	Average recall across classes, reducing majority-class bias.	$\frac{1}{4}\sum_{k=1}^{4}\frac{TP_k}{TP_k + FN_k}$
Macro F1-Score	Class-weighted harmonic mean of precision and recall.	$\frac{1}{4}\sum_{k=1}^{4} 2 \cdot \frac{\Pr e\, cision_k \cdot Recall_k}{\Pr e\, cision_k + Recall_k}$

Table 1 Model evaluation metrics and corresponding formula

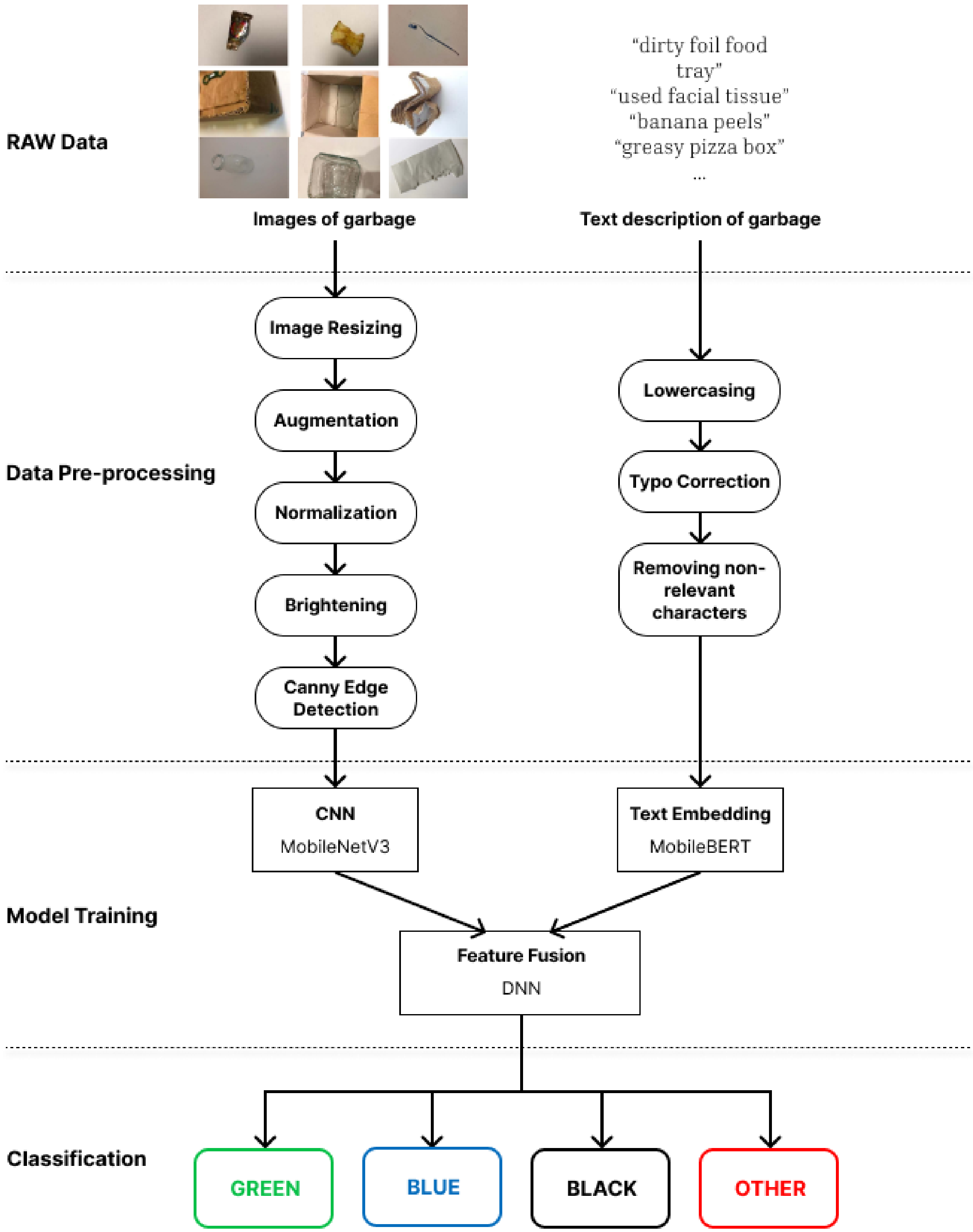

Figure 1 Multimodal Garbage Classification Pipeline

References

[1] City of Calgary. (n.d.). *What goes where?*. https://www.calgary.ca. https://www.calgary.ca/waste/what-goes-where/default.html

[2] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1314-1324).

[3] Yang, Z., Xia, Z., Yang, G., & Lv, Y. (2022). A Garbage Classification Method Based on a Small Convolution Neural Network. Sustainability, 14(22), 14735. https://doi.org/10.3390/su142214735

[4] Sun, Z., Yu, H., Song, X., Liu, R., Yang, Y., & Zhou, D. (2020). Mobilebert: a compact task-agnostic bert for resource-limited devices. *arXiv preprint arXiv:2004.02984*.

[5] Wang, C., Qin, J., Qu, C., Ran, X., Liu, C., & Chen, B. (2021). A smart municipal waste management system based on deep-learning and Internet of Things. *Waste Management*, *135*, 20-29.

[6] Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for Multi-Class Classification: An Overview. *arXiv preprint arXiv:2008.05756.*