Multimodal packaging waste brand identification approach for Extended Producer Responsibility traceability

Abstract

Extended Producer Responsibility (EPR) policies in packaging wastes are challenging due to waste traceability in their post-consumer stage. Tracking packages after disposal involves identifying their producers under extreme conditions. Several Computer Vision (CV) approaches for waste material recognition have been successfully tested. However, the identification of waste producers remains unexplored mainly due to difficult conditions for brand recognition and the requirement of large datasets that vary from place to place and over time. We propose a multimodal approach for waste brand identification that utilizes only one "real" image per product for each brand, achieving a macro F1-score of 0.75 with 23 brands and 38 products. The approach leverages package texts and visual features extracted with pre-trained models and predicts the brand using a KNN model with a custom distance based on the Levenshtein distance. Our method employs data augmentation and random word sampling to create synthetic samples from each product image. The KNN model uses random words and a vector of visual features extracted with a previously trained CNN model for prediction. During prediction, the distance of the K nearest neighbors is computed as the weighted sum of the L^2 visual features distance and the sum of the minimum words Levenshtein distances. This study demonstrates the feasibility of brand identification on packaging waste for EPR traceability without the burden of large dataset acquisition.

Keywords: Extended Producer Responsibility, Multimodal classification, Waste Management, Brand identification, One-shot classification, Machine Learning

---

1. Introduction

Global waste generation is projected to reach 2.59 billion tons annually by 2030, with expectations soaring to 3.40 billion tons by 2050. This represents a substantial increase from the 2.01 billion tons recorded in 2016. On a global scale, the predominant waste category is food and green waste, comprising 44% of total waste. Dry recyclables, including plastic, paper and cardboard, metal, and glass, constitute an additional 38% of waste. The remaining 18% is distributed among rubber and leather, wood, and other (Kaza et al., 2018). Packaging waste, especially plastics, harms the environment and human health when not recycled or disposed of correctly (Li et al., 2021). Given the natural resistance of plastics to degradation, plastic particles may persist in the environment for extended periods, resulting in physical, chemical, and biological harm to organisms (Li et al., 2021).

The OECD (Organisation for Economic Co-operation and Development) defined the EPR (Extended Producer Responsibility) as “an environmental policy approach in which a producer’s responsibility for a product is extended to the post-consumer stage of a product’s life cycle” (OECD, 2016). The core idea is that producers and sellers have some responsibility for the products’ end-of-life environmental impact. The EPR has two main objectives: (i) to shift responsibility from municipalities to producers by holding them accountable for collecting and sorting end-of-life products, individually or collectively, and (ii) to provide incentives for designing products and packaging that facilitate post-consumption management. Several studies examine the implementation and effectiveness of EPR policies in managing packaging waste and promoting traceability to achieve sustainability goals. For instance, Bassi *et al.* in (Bassi et al., 2020) highlight the challenges and potential of EPR in managing plastic packaging waste, emphasizing the need to maximize collection rates and minimize impurities in recycling. Additionally, they identify economic sustainability issues recyclers face, underscore the importance of improving recyclability, and explore the market for secondary plastic.

Daoud and Trigui (Daoud and Trigui, 2019) indicated that one of the challenges in traceability systems is storing and transferring consumer data. Smart packaging offers consumers additional information, such as production methods, promotion, website links, videos, transport tracking, and certification labels. EPR policies can be implemented through different instruments that impact waste generation, product/packaging design, and virgin raw material use. For instance, these policies may include product take-back mandates and recycling rate targets; product take-back mandate and recycling rate targets combined with a tradable recycling credit scheme; voluntary product take-back with recycling rate targets; advance recycling fees (ARF); ARF combined with a recycling subsidy (Walls, 2006). EPR-based policies where the producer is held accountable for its end-of-life products are referred to as "individual" EPR.

In practice, in most cases, the implemented policies correspond to "collective" systems managed by third-party Producer Responsibility Organizations (PRO). Collective systems are easier for the government to supervise than tracking individual companies, allowing for economies of scale in waste collection and risk sharing. On the other hand, individual systems incentivize companies to develop eco-friendly products as each company is uniquely responsible for its product waste management (Walls, 2006). However, with the current technological infrastructure, it is challenging to implement product distinction for some kinds of products, especially for single-use packaging, as it would be required to identify the brand of the disposed packages under conditions such as extreme deformations, contamination, or high degree of occlusions. ERP aims to encourage companies to take back their packaging, leading to a considerable increase in recycling rates in countries with these strong policies. One of the most classic and successful ERP mechanisms is the Deposit Return System (DRS). In this system, each beverage package has a cash deposit when the user buys the product. Later, when the user returns the package, they receive their deposit.

Computer Vision (CV) solutions are a growing research field in waste segregation. In CV-based waste segregation systems, cameras capture images of waste, which are then analyzed by CV algorithms to identify the different types of materials. The CV-based waste segregation is performed either on the generation source or in a centralized place (Lu and Chen, 2022a). Despite promising results, waste identification through CV faces challenges due to real-world complexities, limited datasets, and errors resulting from visual similarities between material packaging (Arbeláez-Estrada et al., 2023). Furthermore, brand identification in waste presents additional complexities; there are often more classes than for material identification, products of the same brand may appear different, product packaging changes over time, and new brands frequently enter and exit the market. In some cases, the elements allowing brand identification consist of small details that can be easily occluded, and brand varieties vary from place to place, making dataset reuse difficult. These particularities require an agile approach to incorporate new classes into CV systems without assembling new datasets comprising multiple variations of product waste. For these reasons and to the best of the authors' knowledge, brand identification in waste has not been investigated for traceability in extended producer responsibility policies.

Therefore, this article proposes a multimodal approach for waste brand identification that leverages package texts and visual features extracted with pre-trained models. It predicts the brand using a KNN model with a custom distance based on the Levenshtein distance (Section 4.2.1). The advantages of the proposed method are that it requires only one sample per product for each brand to be identified, minimal model training is needed, and brands can be easily added, updated, or removed by creating synthetic samples. The main contributions of this article are as follows:

i Construction of a dataset for brand identification in single-use packaging waste (Section 3.2).

ii Proposal of an efficient data preparation pipeline for hyperparameter exploration (Section 4.1).

iii Evaluation of three commonly used approaches for multimodal brand identification (Section 4.2).

iv Proposal of an approach for brand identification in waste that can be trained using only a single image per product from each brand (Section 4.2.1).

v Conduction of ablation and deployment studies of the proposed approach (Section 4.3).

The article is structured as follows: Section 2 presents the state-of-art related to waste packaging traceability and brand identification using CV. The proposed work follows the CRISP-DM methodology, and the results of each stage are presented in Section 3 (Problem contextualization) and Section 4 (Experimentation). Finally, conclusions are drawn in Section 5.

2. Related work

This section explores two key areas within the field of waste management through the application of computer vision (CV) techniques: waste detection and classification and brand or logo detection. The former focuses on developing algorithms to identify and categorize various waste materials, while the latter addresses the challenge of identifying brands or logos within waste streams. These areas are essential for the proposed method outlined in this article, which focuses on efficiently identifying brands through advanced computer vision techniques.

2.1. Waste detection and classification

Different computer vision (CV) approaches in waste materials recognition have been proposed in recent years (Lu and Chen, 2022b). These proposals mainly aim to use simplified environments, artificially collected data, and CV algorithms that can consider the complexities of real-world scenarios related to industrial waste classification. Other proposals present experiments using datasets to establish a set of algorithms for waste detection. These algorithms aid in generating reference datasets that simplify the detection and classification of waste into possible waste categories such as bio, glass, metal and plastic, non-recyclable, cardboard/paper, and other unknown items (Majchrowska et al., 2022).

For instance, Liang and Gu (Liang and Gu, 2021) propose a multi-task learning framework based on a convolutional neural network to recognize and locate wastes in images. Bobulski and Kubanek (Bobulski and Kubanek, 2021) use deep learning to automatically separate plastic waste into four categories: PS (polystyrene), PP (polypropylene), PE-HD (high-density-polyethylene), and PET (polyethylene-terephthalate). The proposed system uses an RGB camera and a microcomputer with CV software. Shengping *et al.* (Wen et al., 2023) propose a deep-learning schema to achieve dynamic and real-time detection of plastic. The schema promotes the quality and efficiency of sorting. It combines the YOLOX (You Only Look Once) object detection model and the DeepSORT (Deep Simple Online and Realtime Tracking) multiple object tracking algorithm. Chu *et al.* (Chu et al., 2018) propose a multilayer hybrid deep-learning system to sort waste automatically. The system integrates a high-resolution camera to capture waste images and sensors to detect other useful feature information. Adedeji and Wang (Adedeji and Wang, 2019) propose a system to classify waste into different types, such as glass, metal, paper, and plastic.

Kumsetty *et al.* (Kumsetty et al., 2023) from another perspective, seek to improve the quality of existing waste datasets using "transfer learning based models such as ResNet and VGG for fast and accurate classification." Training, validation, and testing activities were conducted using TrashNet and TACO datasets. The accuracy achieved on TrashNet was 93.13% and 16% on TACO.

Other authors combine deep learning with the Internet of Things (IoT). For example, Wang *et al.* Wang et al. (2021) propose a waste management system that uses the deep learning-based classifier and cloud computing technique to achieve high-accuracy waste classification at the beginning of garbage collection. Recyclable waste is divided into plastic, glass, paper or cardboard, metal, fabric, and other recyclable waste. Rahman *et al.* Rahman et al. (2022) utilize a microcontroller with multiple sensors, enabling control of real-time data from anywhere through an Android application. Sheng *et al.* Sheng et al. (2020) propose a waste management system by implementing sensors, LoRa communication protocol, and TensorFlow-based object detection. The bin has several compartments to segregate the waste, including metal, plastic, paper, and general waste.

Some works focus on applying machine learning and computer vision techniques to address various aspects of waste management and environmental issues. For instance, Ramirez *et al.* (Ramirez et al., 2020) present a one-shot learning-based classification for segregating plastic waste. Zhang *et al.* (Zhang et al., 2013) present a multi-resolution strategy effective for extracting the open-air informal municipal solid waste dumps. This article focuses on one-shot learning for classifying plastic waste, contributing to effective waste segregation. Finally, Iordache *et al.* (Iordache et al., 2022) concentrated on aerial surveillance stands out as a valuable alternative among methods used to spot littered areas.

2.2. Brand or logo detection

Waste producer identification still needs to be explored mainly due to difficult conditions for brand recognition and the need for large datasets that vary from location to location and over time. Brand detection is a specialization of object detection. It uses logos in different fields to identify trademarks in images or videos. Some AI platforms and research proposals address different approaches to discover which brands are most popular on social networks or most frequent in presenting products in the media, intelligent transportation, and video advertising recommendation (Hou et al., 2023a).

Detection of company logos is significant in object detection. However, Eggert *et al.* (Eggert et al., 2017) mentioned, company logos often appear incidentally in images rather than being the intended subjects. Consequently, they typically occupy a relatively small portion of the image. Thus, Eggert *et al.* conducted a theoretical analysis and derived a correlation between feature map resolution and the minimum detectable object size under the assumption of a perfect classifier (Eggert et al., 2017). In trademark compliance, Chen *et al.* (Chen et al., 2021) propose a robust and highly optimized logo detector that includes general object detection and data augmentation. The logo detection model was achieved from 515 categories in e-commerce images and includes features such as long-tail distribution, small objects, and different types of noise.

Bombonato *et al.* (Bombonato et al., 2018) propose a real-time brand logo recognition system. They experimented with different approaches based on the Single Shot MultiBox Detector (SSD). They used data augmentation and transfer learning to surpass the lower data issue and allow deeper networks. Bianco *et al.* (Bianco et al., 2017) propose a method for logo recognition using deep learning. They evaluated the effect on synthetic versus real data augmentation recognition performance and image pre-processing.

Due to the high degree of unmonitored markets on social networks and other media that imply ubiquity, there is a phenomenon of unauthorized use of brands, especially their graphic images. Trappey *et al.* (Trappey et al., 2022) propose an intelligent system that integrates two models to detect, locate, and crop logos published online as images from product views or displayed on human models. The first model is responsible for logo detection and localization and crops similar trademark-like images from complex online product photos. The second model uses Yolo v4 to locate every cropped logo image and compares them with classes of registered trademarks.

Sujuan Hou *et al.* (Hou et al., 2023b) review advances in logo application using Deep Learning across different fields. The main challenges identified are (i) Small-sized logos, (ii) Images with diverse backgrounds, and (iii) Sub-branding, where products under the same umbrella brand have similar appearances. The future research directions include lightweight detection models, approaches for partially labeled data, video logo detection, tiny logo detection, methods for dealing with long-tail class distributions, and incremental addition of new brands.

Few-shot detection is a machine-learning technique that addresses scenarios with limited labeled training data. Sujuan Hou *et al.* (Hou et al., 2023c) employ a traditional two-stage object detection model and propose a detection head for few-shot logo detection. Training is performed in two stages: the base model is trained with abundant data, and the fine-tuning model is trained with novel balanced k-shots classes. Mikhail Shulgin *et al.* (Shulgin and Makarov, 2023) propose a two-step, zero-shot framework. The first step involves training a universal YOLOv4 logo detector with a large dataset, and the second step uses a pre-trained CLIP model to classify the detected region in a zero-shot manner. Similarly, Mikhail Ermakov and Ilya Makarov (Ermakov and Makarov, 2022) also use a two-stage approach with a universal YOLOv5 logo detection model. The second stage utilizes a pre-trained feature extractor ensemble and a few-shot fine-tuned head to predict the brand.

To the best of the authors' knowledge, producer identification through brand identification in packaging waste has not been investigated. Although the identification of logos and brands has been previously explored, brand identification in packaging waste presents additional challenges, such as high deformations, self-occlusions, and a market where the graphical appearance of packaging constantly changes. Furthermore, we propose a novel approach: (i) using only one image per class for training and (ii) combining the text and graphic elements of the packaging to perform brand prediction. Additionally, this study performs a sensitivity analysis of its hyperparameters and a serverside deployment analysis.

3. Problem contextualization

For this work, we used the CRISP-DM methodology, a process model for carrying out data mining projects (Wirth and Hipp, 2000). It includes the following steps: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. This section covers the phases of business understanding and data understanding. Data preparation, modeling, evaluation, and deployment phases are detailed in section 4.

3.1. Business understanding: Extended Producer Responsibility and policies

Performing proper waste separation at the generation site, before waste transportation, is a critical step for recycling (Alalouch et al., 2021). However, waste separation is challenging for citizens as it is influenced by multiple factors, such as physical and socio-economic barriers, human behaviors, and other reasons (Oluwadipe et al., 2021). Therefore, automatic systems have been proposed to assist with waste separation. The separation process can be performed in two locations: (i) where the user directly disposes of their waste, where waste detection is usually conducted in a chamber (Longo et al., 2021), or (ii) in a centralized location, where the waste is sensed while being transported on a conveyor belt (Mahat et al., 2018). In either of these two systems, the proposed method can be added to extract waste producer information, as most automatic separation systems use RGB cameras and pre-trained feature extractors (Arbeláez-Estrada et al., 2023).

Extended Producer Responsibility (EPR) schemes have gained widespread adoption in recent years, particularly in Europe and other regions, significantly increasing material and energy recovery from waste streams (Dalhammar et al., 2021). Researchers have increasingly recognized EPR as a potential solution to the global plastic pollution problem, with studies focusing on its implementation and effectiveness in improving plastic waste management practices, particularly in densely populated regions like the European Union (EU) (Lorang et al., 2022).

Pouikli (Pouikli, 2020) highlights several benefits of EPR: (i) The creation of more efficient separate collection schemes for specific waste streams. (ii) The minimization of the burden on public budgets by shifting financial responsibility for products' end-of-life phases from local municipalities and public authorities to producers. (iii) The generation of separated, high-quality waste materials supports the development of secondary raw materials markets. (iv) The encouragement of producers to move towards eco-design innovations to reduce waste management costs. And (v) the promotion of technological and organizational progress and contribution to resource security by diversifying the material supply sources. Pouikli also mentions weaknesses in implementing existing EPR schemes: (i) the lack of a harmonized definition and scope for EPR. (ii) There is no transparent information and fragmentation regarding cost coverage. (iii) The limited influence of EPR schemes on eco-design improvements. (iv) The inadequate control and monitoring mechanisms. And (v) the failure to accurately determine the number of costs that should be internalized through recycling targets.

The OECD (OECD, 2016) categorizes EPR instruments into four main types: (i) Product take-back requirements, which involve assigning responsibility to producers or retailers for end-of-life management of products. (ii) Economic and market-based instruments provide a financial incentive to producers to implement the EPR policy through several forms, including deposit refunds, advance disposal fees, material taxes, and upstream combination taxes/subsidies. (iii) Regulations and performance standards, including technical standards and minimum mandatory recycling rates. And (iv) information-based instruments aim to indirectly support EPR programs by raising public awareness via reporting requirements, labeling of products, and information campaigns for consumers about producer responsibility and waste separation.

Some authors have reviewed and analyzed EPR cases in specific locations. For instance, Gupt and Sahay in Gupt and Sahay (2015) review 27 cases of EPR from developed and developing economies to ascertain its most important aspect. The results indicate that in developed countries, producers carry a higher financial responsibility with little physical responsibility for recycling. Conversely, in developing countries, producers are more evenly tasked with financial and physical responsibility. EPR is very successful in developed economies, moderately successful in developing economies without an informal sector, and unsuccessful in developing countries with an informal sector. Particularly, the Colombian EPR focuses on Waste from Electrical and Electronic Equipment (WEEE), such as batteries, bulbs, and computers. Consumers are responsible for separating the products from municipal solid waste and bringing them to the retailers. Retailers accept the used products from the consumers for free and act as producer collection points. The scheme sets collection targets, including yearly increases and a medium-term target. There has been significant improvement in the collection rate, while the recovery continues to be limited.

Nonetheless, the law only requires producers to assume responsibility for managing the end-of-life of packaging waste. Therefore, the actions or decision-making might be biased in how their brand moves through sorting. One alternative to improve that situation can be an intelligence classification model for packaging waste throughout its life cycle, which might drive another kind of decision-making.

Governments and companies worldwide are including programs and laws in their strategic agendas to achieve Extended Producer Responsibility (EPR) (Watkins et al., 2017). It means that product manufacturers, importers, and brands assume financial responsibility and, in some cases, physical liability for the environmental impacts of their products throughout their product life cycle. Specifically, the EPR for containers and packaging requires producers to assume responsibilities for managing the end-of-life of their product containers. Recently, (Somlai et al., 2023) presented an analysis of the Member States of the EU statistical reports on the generation of plastic packaging waste. It starts by exploring the quality of the reports based on the two approaches used to calculate the generation of packaging waste: placed on the market and waste analysis. The findings revealed that EU members have different statistical approaches using other variables, leading to different results. Factors such as parasitism, non-compliance, and insignificance can be detected, which cause weaknesses in the evaluation and, consequently, distortion in the presentation of statistics on packaging waste. This behavior is because the producers have financial incentives to declare less than necessary.

Baxter *et al.* (Baxter et al., 2022) analyze brand information and beach cleanup data from five locations in Canada to determine the efficacy of ongoing single-use plastic (SUP) mitigation measures. Litter was collected, sorted, categorized, and recorded into the categories of brand, product description, number of items collected, product use, and type of plastic. The results show that "six prevalent litter brands (Nestlé, PepsiCo, Coca-Cola, Tim Horton's, Starbucks, and McDonald's) comprise 45% of known branded litter collected for urban study locations and comprise 39% of branded litter collected in all study locations" and "that current Canadian SUP mitigation measures are likely insufficient to adequately reduce SUP leakage into natural environments."

Stanton *et al.* (Stanton et al., 2022) conducted a study in the United Kingdom based on citizen science on Anthropogenic Litter (AL - the one humans produce with our activity). The study identified vital materials, industries, brands, and parent companies associated with AL. The findings showed 63% plastic, 14% metal, and 12% composite materials. Most AL

(56%) were used as beverage and non-beverage containers. Of the branded AL, 26% was associated with The Coca-Cola Company, Anheuser-Busch In-Bev, and PepsiCo.

Related to policies and regulations, Tumu *et al.* (Tumu et al., 2023) conducted a review of global EPR and recycling laws, considering regulations from the United States, European Union, and UK. The study found that countries with established plastic regulations and landfill bans have higher recycling rates. Similarly, Ya-Jun Cai *et al.* (Cai and Choi, 2019) performed a systematic review to identify innovative proposals related to EPR in five areas: policies, product design, process, supply chain, and technology. From policies, they identified three key elements for enhancing sustainability: (i) stopping illegal and informal recycling, (ii) forming alliances between countries, and (iii) evaluating policies with quantitative data. Additionally, several studies have focused on specific countries' policies, such as those in China (Zhu et al., 2019), Germany and the UK (Ramasubramanian et al., 2023), Brazil (de Miranda Ribeiro and Kruglianskas, 2020), the USA (Nash and Bosso, 2013), and Colombia (Park et al., 2018), among others.

3.2. Data understanding: Packaging waste brand dataset

This study utilizes two datasets, named SRC (Source) and DST (Destination), developed by the research team to analyze single-use food packaging waste from locally consumed brands (available in the BrandWaste repository[1]). Each sample in the dataset consists of an image of the waste, a list of visible words from the packaging captured in the image, and its corresponding label, which is the brand intended to be predicted by the model.

The SRC dataset comprises images of products taken before consumption, capturing all sides of the text, and is employed to train the models. Similarly, the DST dataset includes images, package texts, and labels but focuses on products after consumption. Various team members capture images for the DST dataset without any specific indication or restriction regarding the photos. Examples of images from the DST and SRC datasets are illustrated in Figure 1.

Figure 1: SRC and DST dataset images with text extraction examples.

The datasets comprise 23 brands, each with various products. For example, there are images from three different products of the brand "Tosh": two types of tea infusions and a cereal bar. Figure 2 displays the number of

[1]https://github.com/juancc/OpendataWasteDatasets.git

images for each brand, with different colors representing the products within each brand in the DST dataset. The DST dataset contains 1008 images and encompasses 38 products solely intended for model evaluation.

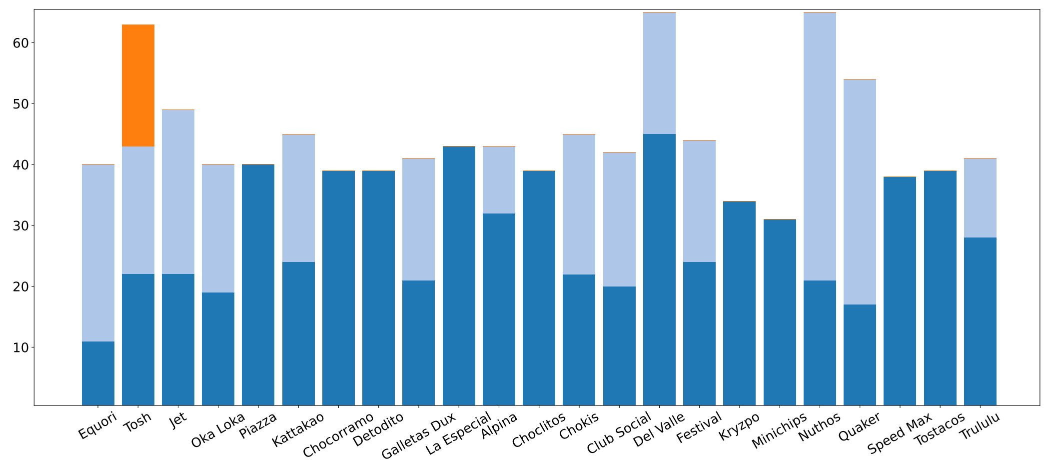

Figure 2: Distribution of images by brands and product references of the DST dataset.

4. Experimentation

4.1. Data preparation

The process of text extraction from the images of both datasets utilizes a CRAFT pre-trained model (Baek et al., 2019) for text detection and a Convolutional Recurrent Neural Network (CRNN) pre-trained model (Zdenek and Caicedo, 2021) for text recognition. The text detection identifies image regions containing text, and the recognition model predicts the text within the detected regions. The text extraction model evaluated with 60 DST images has an average of 4.1 missing words, 2.1 average Levenshtein distance, and three extra words. Most errors occur in texts with a vertical layout in the text extraction model. A primary limitation of using a pre-trained recognition model is the presence of noisy text predictions, as it may not be tuned to the fonts or language used in the packages. Additionally, due to the nature of the problem, it is not feasible to train or fine-tune the text extraction model within a highly dynamic product market with frequent rotations. Thus, the proposed approach is robust to small variations in the same words, making it suitable for a pre-trained text recognition model.

Figure 3 illustrates the distribution of words per class. The DST dataset has an average of 55.5 words per class (dashed gray) with a Standard Deviation (SD) of $\sigma = 66.4$, while SRC has an average of 482.1 words per class (dashed red). SRC contains almost seven times more words per class on average than DST because SRC includes all the packages from all views, while DST's texts are a subset of SRC. Consequently, most DST packages are highly deformed, and the words are self-occluded. Additionally, there is significant variation in the number of words per class, particularly noticeable in SRC. The longest class has 1173 words, and the shortest has 185 words. The variability in the number of words is highly dependent on the brand, primarily due to ingredient variations and importation information.

The text extraction process for both SRC and DST is identical, except that in SRC, all product views from all references within one brand are merged into one sample. Consequently, in our case, the number of samples in SRC is 23. Additionally, an image without its background is added to the SRC dataset from the online catalog for each product reference. The data

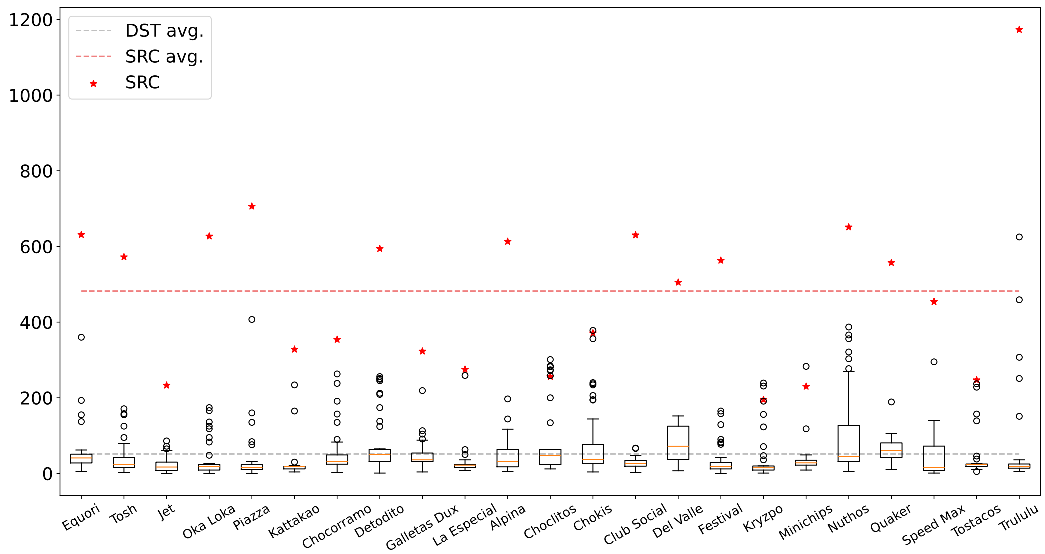

Figure 3: The number of words in the SRC dataset is depicted in red, while the box plot illustrates the distribution of the number of words in the DST dataset. On average, the DST dataset contains 55.5 words per class (dashed gray), with an SD of $\sigma = 66.4$. In contrast, the SRC dataset averages 482.1 words per class (dashed red).

augmentation and feature extraction discussed in the next section utilize these catalog images.

The pipeline for constructing the test and train datasets consists of three steps: (i) data augmentation, (ii) feature extraction, and (iii) distance caching. This construction pipeline takes the described SRC and DST datasets as input. It returns the train and test datasets used in the modeling and evaluation stages of the CRISP-DM methodology.

4.1.1. Data augmentation (DA)

The training dataset consists of synthetically generated copies of the SRC samples. The process involves image transformation and random text sampling. The applied image transformations include:

1. Similarity transformations: rotation, translation, and scaling.
2. Image blurring with a box filter of $(4, 4)$ size and 0.5 probability for each image to be blurred.
3. Vertical and horizontal image flipping.
4. Image contrast adjustment: point-wise operation per image channel given by $r(I_c - \bar{I}_c) + \bar{I}_c$, where r is the random contrast factor, I_c represents the intensities in channel c, and $\bar{I}_c$ is the mean of the channel.
5. Random background insertion: the resulting image C is obtained through the alpha matting operation: $C = (1 - \alpha)B + \alpha F$, where B is a random background chosen from either an image selected from (Lprdosmil, 2022), Gaussian noise, or flat random color. The alpha channel α is generated from the catalog image F. Similarity transformations are also applied to the background images, and this augmentation is applied to all images.
6. Noise addition: zero-centered Gaussian noise with $\sigma = 15$ is added to the images with a 0.5 probability of occurrence.

The transformations used to generate the synthetic copies aim to create viable variations due to waste position relative to the camera (transformation 1), camera-associated distortions (transformations 2 and 6), lighting changes (transformation 4), and adding variability to the dataset (transformations 3 and 5). The parameters used in the transformations are defined based on the best results achieved in the exploratory tests.

The text augmentations consist of random sampling from SRC texts to form a fixed-length list of 200 words per observation. The number of words selected is based on approximating the SRC class with the fewest words. Both image and text augmentations occur offline to reduce training time. Therefore, the training dataset comprises synthetically generated random copies of the SRC dataset.

4.1.2. Feature extraction

The image features correspond to a 2048-length vector composed by the global average pooling of the last convolutional layer of a ResNet50 model pre-trained on ImageNet (He et al., 2016). In this process, an image is passed through several convolutional blocks. Finally, the average is computed along the channel axis of the feature map volume returned by the last convolutional block. These features are extracted from the synthetic images created with the data augmentation (DA) and stacked together in a matrix for the next step (Section 4.1.3). This study refers to them as CNN (Convolutional Neural Network) features.

Text extraction features only apply to the Bag-of-Words (BoW) Neural Network (NN) model (Table 1). A vocabulary is generated consisting of the unique words found in the texts of all the brands from SRC. A word is added to the vocabulary if its length is greater than 3 and it is not a number. The vocabulary is then used to extract features that describe a text. Thus, the text features of a sample comprise a vector of size equal to the vocabulary, containing the number of times (frequency) each word from the vocabulary appears in the text. In order to account for slight variations between two words that are essentially the same but may differ due to text extraction errors, plurality, or gender variations, the Levenshtein distance is used.

The Levenshtein distance (Levenshtein et al., 1966) is the minimum number of operations (insertions, deletions, and replacements) needed to transform a word V into W at their corresponding character positions a and b, denoted as $d_L(V,W)$. Equation 1 represents the Levenshtein distance between V and W, and it can be computed using dynamic programming (Wagner and Fischer, 1974).

$$d_L(aV,bW) = \begin{cases} d_L(V,W) & \text{if } a = b \\ 1 + \min\left(d_L(V,W), d_L(aV,W), d_L(V,bW)\right) & \text{otherwise} \end{cases} \tag{1}$$

Therefore, the text feature vector C of a synthetic sample used in BoW models is the distribution of vocabulary words $V = \{v_0, \ldots, v_j\}$ in the random words of the synthetic sample $W = \{w_0, \ldots, w_i\}$. To compensate for slight variations in words, the Levenshtein distance is used. If the distance exceeds the threshold value t, a $+1$ is added at the corresponding position j

of the vocabulary list, as shown in Equation 2.

$$C(j) = \begin{cases} +1 & \text{if } d_L(W_i, V_j) > t \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The feature extraction takes place offline, resulting in a matrix with a height equal to the number of synthetic copies and a width equal to the length of the concatenated features.

4.1.3. Distance caching

In order to minimize the time required to perform the ablation studies, the distances between the DST dataset and the training dataset are precalculated. Two matrices are computed: (i) a visual features distance matrix, which represents the L_2 distance between the 2048-dimensional vectors extracted with the ResNet50 model, and (ii) the text distance matrix, which is composed of distances d between the two lists of words ($l^{(1)}$, $l^{(2)}$) of training and DST samples. The distance is computed as the sum of the minimal Levenshtein distances between each word of $l^{(1)}$ and $l^{(2)}$ (Equation 3). The median of the CNN distances is 46, and the text distance is 54.

$$d\left(l^{(1)}, l^{(2)}\right) = \sum_{j}^{m} \min \sum_{i}^{n} d_L\left(l_j^{(1)}, l_i^{(2)}\right) \tag{3}$$

4.2. Modeling: Wastes brand identification strategy

The approach taken to modeling is to first test different Machine Learning (ML) techniques in a simplified case of predicting ten brands. Later, ablation studies are performed on the most promising ML techniques. Therefore, this section describes the initial model approaches explored in this study. Table 1 presents the evaluation results on the DST dataset of each model alternative, ordered by F1-score macro. The F1-score is the harmonic mean between precision and recall by class, and the macro F1-score is their average value. The exploration tests perform multiple model configurations with different hyperparameters; the best ones are reported. The term "encoding" refers to transforming images and texts into numeric values used as input for the models. The different encodings correspond to three different scenarios: using only image features, using only textual features, and using both. The exploratory tests consider two ML techniques (see Table 1): feedforward Neural Network (NN) and K-Nearest Neighbors (KNN). Each approach is identified by its respective ID. The exploration tests involve using different encoding types to convert the texts and images into suitable forms for training the models. Thus, both model alternatives take as inputs the features and distance matrices described in the previous Section 4.1. The performance of each model is compared by evaluating the macro F1-score on the DST dataset.

Table 1: Initial modeling approaches exploration evaluated in DST with 10 brands.

ID	**Architecture**	**Encoding**	**F1-score**	**Acc.**
K-LC	KNN	Levenshtein, CNN	0.81	0.81
K-L	KNN	Levenshtein	0.72	0.72
N-BC	NN	BoW, CNN	0.66	0.69
N-W	NN	Words Embedding	0.61	0.61
N-C	NN	CNN	0.58	0.61

4.2.1. KNN

KNN predicts the brand of a DST observation by finding the *mode* of the brand of the K nearest synthetic observations. The brute-force algorithm is used to find the nearest samples. Using the pre-calculated distance matrices, this is achieved by sorting each row by the lowest value and finding the *mode* of the first K corresponding labels. The data encoding for the KNN model utilizes the Levenshtein-based distance for text encoding and the CNN feature extraction for image encoding, as described in Section 4.1.3.

The KNN tests explored the results of using a weighted combination of the two distance matrices, such that the resulting distance is $D = \alpha A + \beta B$, where A is the visual features distance and B is the distance of the texts. The α and β control the balance between text and visual features. Thus, $\beta = 0$ means only visual information is used for prediction. In our preliminary test, we used only visual information, which led to poor model performance. Therefore, for the rest of the experiments, we set $\beta = 1$ and use α to tune the balance of features.

Table 1 presents the results of the exploratory tests with $\alpha = 0$ (K-L) and $\alpha = 3$ (K-LC), using a training dataset composed of 4000 synthetic samples (400 random copies of each class) and $K = 5$. It is worth highlighting that the text-only KNN achieves a higher F1 score than the visual model alone, but when they are combined, the best models are achieved by using higher α values.

4.2.2. NN

The preliminary tests explored three NN variants, each consisting of a two-dense-layer NN model (head) on top of three different encoding methods (see Figure 4). The variants also include a *Dropout* operation with a percentage of 0.5 applied to their input features to mitigate the overfitting generated by using only one sample per class for training. Figure 4 shows the diagram of the explored NN model variations, where the three variants use different input data as commonly employed. The subsequent section describes each explored variant in detail. The exploratory tests considered different hyper-parameters for each variant and reported the ones with the highest scores on the DST dataset (Table 1).

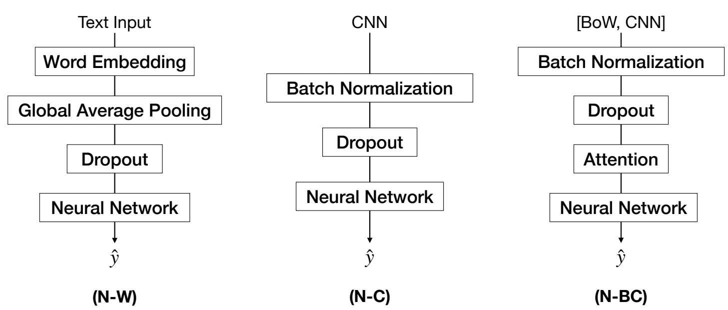

Figure 4: Diagram of three NN variants architectures explored during initial tests of predicting 10 brands. Table 1 presents the results in DST dataset.

The N-BC model (Figure 4) utilizes the concatenation of the CNN and BoW features (Section 4.1.2) to simultaneously use vector-encoded visual and textual features. The model includes an *Attention* layer that performs an element-wise multiplication between a mask and the model's inputs. A dense layer with Sigmoid activation learns the mask values. The first dense layer of the model's head has 128 units, and the model achieved an accuracy of 0.69 and an F1-macro score of 0.66 on the DST dataset. This approach allows the model to learn connections between specific words and image patterns directly and determine which connections are more relevant for predicting the brand.

N-W is the next best-performing model on the DST dataset (accuracy of 0.61 and F1-macro score of 0.61). This NN variant uses *word embedding* encoding, where a fixed-length representation of the input words is produced by learning encoding parameters during the model's training. The channel-wise average of the densely encoded words feeds the NN model's head. The N-W model utilizes a vocabulary of 5000 words with 28 embedding dimensions and a head with 128 intermediate dense units. Using a word embedding approach, the model learns text encoding during training, in contrast with the other types of text encoding used in this study (BoW and Levenshtein-based).

The last variant, N-C, follows the *traditional approach* for image classification with Deep Learning. This approach involves placing a custom head on top of an ImageNet pre-trained feature extractor. The best configuration achieved for this approach includes a training batch size of 8, a learning rate of 0.001, and 256 intermediate dense head units (accuracy of 0.61 and F1-macro score of 0.58).

The training of the three variants uses the Adam optimizer (Kingma and Ba, 2014), with *categorical cross-entropy* loss and 0.1 label smoothing (Müller et al., 2019). Additionally, 10% of the training dataset is used for validation split.

4.3. Evaluation

Based on the exploratory tests, the most promising model is KNN because it achieved the best score and possesses additional properties that align with the project requirements. The KNN model does not require intense training like the other alternatives, and it is easy to add, update, or remove a brand simply by modifying the synthetic copies of the model. Therefore, the next sections present a further examination of the KNN model.

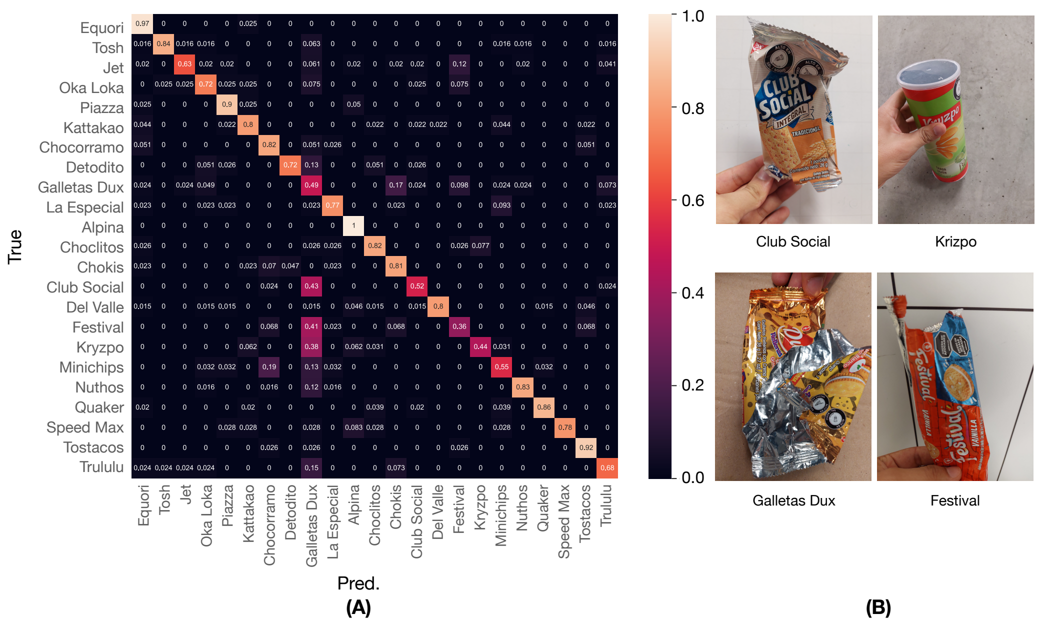

Figure 5: In **A**: Confusion matrix normalized by row showing the model performance on the DST dataset with 23 classes. The y-axis represents the true labels, and the x-axis represents the model's predictions. Model hyperparameters are $k = 17$, SRC model samples: 11000, and $\alpha = 5$. In **B**: Examples of images of brands with lower performance.

The model is evaluated on the entire DST dataset, which comprises 23 brands (Figure 2), using the macro F1-score metric. Figure 5 shows, in (A), the confusion matrix of the model evaluation on the DST dataset normalized by row. The class with the worst results is "Galletas Dux", which has an F1-score of 0.27. It was mostly confused with three other brands: "Club Social", "Festival", and "Kryzpo". This could be due to their packaging texts being similar, as indicated by our proposed distance metric (Figure 8, lower distances), and some of them visually resembling each other (see part B in Figure 5).

The evaluation shows balanced results in terms of average precision (0.78) and recall (0.74) and, consequently, in the F1-scores (0.75) in the DST dataset with 23 classes. Table A.2 in the Appendix presents the detailed evaluation report per class.

In order to understand the model's behavior, this study performed an exploration of the model's hyperparameters. Grid search was used as the exploration method with 3 hyperparameters for the models: (i) K, the number of nearest neighbors used for prediction, (ii) *samples*, the number of synthetic samples composing the KNN model, and (iii) alpha (α), which controls the balance between visual and textual features (Section 4.2.1). Figure B.9 in Appendix shows the hyperparameter exploration process.

The results of the hyperparameter exploration show that concerning the α value, relying too heavily on either visual or textual features is detrimental to the model performance, but over-relying on textual features has a greater negative impact. On the other hand, K and *samples* are related to having more samples, allowing for more possible variations. With a higher K, more samples are considered for making predictions. In our case, the optimal zone is around 11000 samples, equivalent to ~ 478 copies per class, with $K = 17$ and $\alpha = 5$.

To the best of the authors' knowledge, the proposed method is the first to apply brand identification to waste. Therefore, Table 1 presents a comparison with commonly used approaches ~~for reference~~. However, to provide context, compared to other few-shot approaches discussed in Section 2.2, our proposal achieved an F1-score of 75 with 25 classes. For comparison, Bhunia et al. (2019) obtained a Mean Average Precision (mAP) of 66.8 on the FlickrsLogos dataset with 12 testing classes; (Hou et al., 2023c) reached 74.4 Average Precision (AP) with 10 novel classes in the FlickrLogos-32 dataset; Ermakov and Makarov (2022) achieved 85.83 average accuracy with a five-shot approach in logo classification on the FlickrLogos-32 dataset; and Liu et al. (2021) achieved 78.6 accuracy in logo classification on clothes that have similar properties to plastic packaging (e.g., deformable) but used the entire training dataset. It is important to note that these results cannot be directly compared, as they involve different vision tasks (e.g., logo detection), evaluation metrics, numbers of classes, and domains. On one hand, detection tasks are more complex because they require predicting the coordinates of the bounding box. On the other hand, waste presents different challenges, such as high appearance variation due to deformation and contamination, the absence of logos—a highly recognizable packaging feature—in most cases, and the similar appearance of classes within the same category.

4.4. Deployment: Ablation studies

The ablation studies consider three tests: a server load test, model behavior with adding new classes, and class separability.

4.4.1. Load server test

The load server test analyzes the prediction time on a "production" server of the KNN model without considering the feature extraction. Table C.3 in Appendix presents the server's technical specifications for load tests. The test involved measuring the model response time with different model sizes (number of synthetic samples). Figure 6 presents, in blue, the time required to predict 200 DST samples on one server core depending on the model size (number of synthetic samples), where the size of each sample represents the number of words. The black dashed line indicates the average prediction time. The average time to predict 10 DST samples in parallel by the server cores is shown in red. The parallel prediction is achieved by dividing the samples to predict the number of server cores.

The median prediction time with a model of 400 copies per class is 28.3s per core. The sample with the shortest execution time was 0.09s, while the sample with the longest execution time was 397.9s. The variation in prediction time for each DST sample is explained by the difference in their number of words. The number of words depends on the total words of the pack-

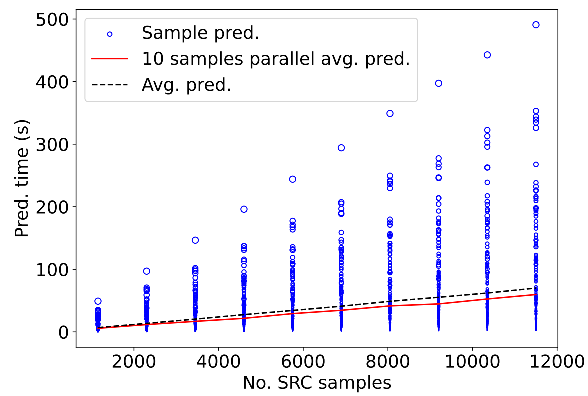

Figure 6: Server time response predicting 200 DST samples. The blue circles represent each DST sample, and their size represents the number of words.

aging and the occlusion due to deformation or photo point of view (Figure 3). This time could be controlled by fixing the number of words the model uses to predict. Nevertheless, this strategy could impact the model's performance. Additionally, two strategies could optimize the prediction time: (i) the dynamic programming of the Levenshtein distance computation directly influences the prediction. This could be improved mainly by parallelizing on the GPU (Castells-Rufas, 2023). (ii) Using non-exact KNN approaches such as KD-trees that split in half the search space, as the proposed method uses many synthetic copies.

4.4.2. Model behavior with the addition of new classes

An important characteristic of brand identification in waste is that brands are continually changing, whether due to the creation or disappearance of brands or updates to the packaging appearance of products. Therefore, this analysis tests the variation of the F1 score by adding new classes.

Figure 7 presents the mean difference by adding new classes to the same hyperparameter model (copies per class: 500, $K = 17$, and $\alpha = 5$). The addition of classes is performed by adding the synthetic samples of the new class to the KNN model. The results show a standard deviation of the F1-score of $\sigma = 0.05$ and a mean F1-score of 0.762. The highest value was reached with 12 classes at 0.84, and the lowest with 19 classes at 0.70. It is worth noting that with the addition of new classes, the model size and the prediction time increase.

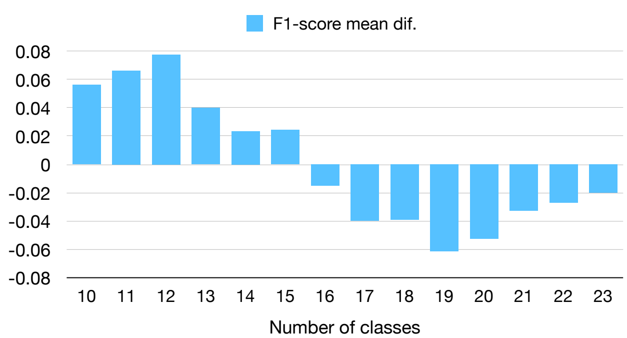

Figure 7: Mean difference in F1-score on the DST dataset by progressively adding classes, starting from 10 classes, while fixing the hyperparameters (copies per class: 500, $K = 17$, and $\alpha = 5$).

The results of the model behavior by adding new classes (Figure 7) suggest no direct link between adding more classes and reducing the model performance. Instead, the performance is affected by the fixed model hyperparameters that need to be tuned for each target class. Although the proposed method uses only one image per class for training, real images are still required for optimal hyperparameter selection and model evaluation. However, compared to traditional DL approaches, the burden of dataset building is minimal, as only evaluation/validation images are required, and there is no need for text labeling, as the proposed approach can use pre-trained text extractor models.

4.4.3. Class separability

The class separability is analyzed by comparing the texts between the brands using the KNN text distance (Equation 3). Figure 8 presents the results of the class separability, where a lower value indicates that, according to the model distance, the texts of the brands are similar.

Because the proposed distance is non-commutative, the distance between two classes is computed as the average of the distances in both directions ($d(A, B)$ and $d(B, A)$) of the texts of classes A and B. The three most difficult classes to differentiate, calculated by the lower value of their column sums, are: "La Especial" (F1-score 0.80), "Minichips" (0.57), and "Galletas Dux" (0.27). The easiest classes to differentiate by the text descriptions are: "Equori" (F1-score 0.84), "Piazza" (0.86), and "Trululu" (0.73). This analysis matches 3 of the least separable classes out of the 5 brands with the lowest F1-scores in the model evaluation described in Section 4.3.

One of the main concerns with using text descriptors for food packaging classification is the potential similarity among texts, as many package texts list ingredients. The analysis of text separability provides insight into the theoretical separability of packages based on their texts and can help identify

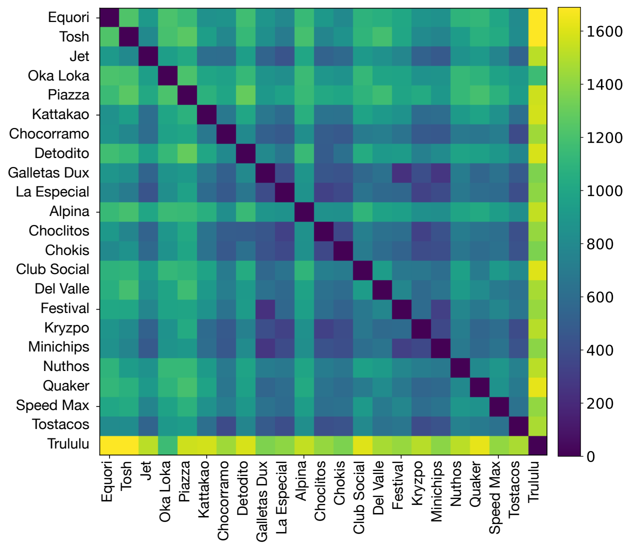

Figure 8: Text distance (Equation 3) between SRC brands. Each brand comprises all the package texts; thus, this matrix shows the theoretical separability between the brands based on their texts.

challenging classes. However, it is important to note that this analysis offers only a partial view of the potential model performance, as our approach utilizes visual and text features.

5. Conclusions

EPR policies are fundamental in encouraging companies to create more environmentally friendly products, and these policies depend on obtaining reliable information throughout the entire product life cycle. However, obtaining the brand of wasted packages is difficult. In this research, an approach for obtaining producer information through images of waste packaging is proposed. This approach can be integrated into automatic waste separation systems already used for recycling. Brand identification is complicated due to the characteristics of waste and the dynamic nature of the product market. While vision-based systems are already being used to classify waste materials, performing waste brand identification poses challenges, primarily due to the large labeled datasets required by current solutions.

We propose a multimodal approach for waste brand identification that relies on only one "real" sample per brand and achieves a macro F1-score of 0.75 with 23 brands and 38 product references (Section 4.3). This approach utilizes package texts and visual features extracted with pre-trained models. It predicts the brand using a KNN model with a custom distance based on the Levenshtein distance (Section 4.2.1).

The proposed method generates synthetic random copies of real samples, which form the basis of the KNN model. Therefore, three hyperparameters control the performance of the model: the number of synthetic samples, the number of nearest neighbors (K), and an alpha (α) value that regulates the balance between visual and textual features. Since the KNN model comprises synthetic samples, the number of copies used to create the model directly impacts prediction time. Similarly, like the other hyperparameters, its "optimal" value depends on the number of brands and their characteristics. However, the most influential factor in prediction time is the number of words in the waste package (Figure 6).

Packaging texts are fundamental descriptors of model performance and can be used to assess brand separability and identify difficult classes in advance by analyzing their distances (Section 4.4.1). Although this only provides a partial view of the model as it only considers textual features, analyzing distances allowed us to identify 3 out of 5 of the worst-performing brands.

Additionally, this study explored Neural Network architectures with 3 text encoding types. However, not only was the KNN model superior in performance, but it also better met the project requirements regarding brand addition, removal, or updating. With the proposed approach, a brand can be added or removed from the KNN model by modifying the synthetic copies of the model. In our experiments of progressively adding 14 classes with the same model hyperparameters, the standard deviation of the F1-score was $\sigma = 0.05$ and a mean F1-score of 0.762.

This study demonstrates the feasibility of brand identification on packaging waste for EPR traceability without the burden of acquiring large datasets. Using only one image per product of each brand and virtually no training, the proposed approach allows for easily adding or updating products and brands. Additionally, this study constructed a dataset for waste brand identification that is publicly available, and evaluated commonly used approaches for brand classification.

The main limitations of the proposed approach are: the large number of synthetic samples per brand needed to achieve good performance, the combinatorial nature of text distance calculation, and that although the proposed approach uses only one "real" sample for training, this study uses labeled target images for model evaluation. These labeled target images would be necessary for model evaluation or cross-validation in a real scenario.

Further work could focus on deeper visual features and text extraction integration with the KNN model. For example, exploring using the same feature extractor for both models could enhance integration. Additionally, the prediction time of the KNN model could be improved by investigating approximate nearest neighbor search alternatives. Moreover, model performance could be enhanced by employing a more accurate text extractor model.

Acknowledgments

This research was funded by Ministerio de Ciencia Tecnología e Innovación of Colombia and the Government of Antioquia by BPIN grant number 2020000100207

References

Adedeji, O. and Wang, Z. (2019). Intelligent waste classification system using deep learning convolutional neural network. *Procedia Manufacturing*, 35:607–612.

Alalouch, C., Piselli, C., and Cappa, F. (2021). *Towards Implementation of Sustainability Concepts in Developing Countries*. Springer.

Arbeláez-Estrada, J. C., Vallejo, P., Aguilar, J., Tabares-Betancur, M. S., Ríos-Zapata, D., Ruiz-Arenas, S., and Rendón-Vélez, E. (2023). A Systematic Literature Review of Waste Identification in Automatic Separation Systems. *Recycling*.

Baek, Y., Lee, B., Han, D., Yun, S., and Lee, H. (2019). Character region awareness for text detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9365–9374.

Bassi, S. A., Boldrin, A., Faraca, G., and Astrup, T. F. (2020). Extended producer responsibility: how to unlock the environmental and economic potential of plastic packaging waste? *Resources, Conservation and Recycling*, 162:105030.

Baxter, L., Lucas, Z., and Walker, T. R. (2022). Evaluating canada's single-use plastic mitigation policies via brand audit and beach cleanup data to reduce plastic pollution. *Marine Pollution Bulletin*, 176:113460.

Bhunia, A. K., Bhunia, A. K., Ghose, S., Das, A., Roy, P. P., and Pal, U. (2019). A deep one-shot network for query-based logo retrieval. *Pattern Recognition*, 96:106965.

Bianco, S., Buzzelli, M., Mazzini, D., and Schettini, R. (2017). Deep learning for logo recognition. *Neurocomputing*, 245:23–30.

Bobulski, J. and Kubanek, M. (2021). Deep learning for plastic waste classification system. *Applied Computational Intelligence and Soft Computing*, 2021:1–7.

Bombonato, L., Camara-Chavez, G., and Silva, P. (2018). Real-time brand logo recognition. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 22nd Iberoamerican Congress, CIARP 2017, Valparaíso, Chile, November 7–10, 2017, Proceedings 22*, pages 111–118. Springer.

Cai, Y.-J. and Choi, T.-M. (2019). Extended producer responsibility: A systematic review and innovative proposals for improving sustainability. *IEEE transactions on engineering management*, 68(1):272–288.

Castells-Rufas, D. (2023). Gpu acceleration of levenshtein distance computation between long strings. *Parallel Computing*, 116:103019.

Chen, H., Li, X., Wang, Z., and Hu, X. (2021). Robust logo detection in e-commerce images by data augmentation. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, page 4789–4793, New York, NY, USA. Association for Computing Machinery.

Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S., Xiong, X., et al. (2018). Multilayer hybrid deep-learning method for waste classification and recycling. *Computational intelligence and neuroscience*, 2018.

Dalhammar, C., Wihlborg, E., Milios, L., Richter, J. L., Svensson-Höglund, S., Russell, J., and Thidell, Å. (2021). Enabling reuse in extended producer responsibility schemes for white goods: legal and organisational conditions for connecting resource flows and actors. *Circular economy and sustainability*, 1(2):671–695.

Daoud, M. K. and Trigui, I. T. (2019). Smart packaging: Consumer's perception and diagnostic of traceability information. In *Digital Economy. Emerging Technologies and Business Innovation: 4th International Conference, ICDEc 2019, Beirut, Lebanon, April 15–18, 2019, Proceedings 4*, pages 352–370. Springer.

de Miranda Ribeiro, F. and Kruglianskas, I. (2020). Critical factors for environmental regulation change management: Evidences from an extended producer responsibility case study. *Journal of Cleaner Production*, 246:119013.

Eggert, C., Zecha, D., Brehm, S., and Lienhart, R. (2017). Improving small object proposals for company logo detection. In *Proceedings of the 2017 ACM on international conference on multimedia retrieval*, pages 167–174.

Ermakov, M. and Makarov, I. (2022). Few-shot logo recognition in the wild. In *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo)*, pages 000393–000398. IEEE.

Gupt, Y. and Sahay, S. (2015). Review of extended producer responsibility: A case study approach. *Waste Management & Research*, 33(7):595–611.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hou, S., Li, J., Min, W., Hou, Q., Zhao, Y., Zheng, Y., and Jiang, S. (2023a). Deep learning for logo detection: A survey. *ACM Trans. Multimedia Comput. Commun. Appl.* Just Accepted.

Hou, S., Li, J., Min, W., Hou, Q., Zhao, Y., Zheng, Y., and Jiang, S. (2023b). Deep learning for logo detection: A survey. *ACM Transactions on Multimedia Computing, Communications and Applications*, 20(3):1–23.

Hou, S., Liu, W., Karim, A., Jia, Z., Jia, W., and Zheng, Y. (2023c). Few-shot logo detection. *IET Computer Vision*, 17(5):586–598.

Iordache, M.-D., De Keukelaere, L., Moelans, R., Landuyt, L., Moshtaghi, M., Corradi, P., and Knaeps, E. (2022). Targeting plastics: machine learning applied to litter detection in aerial multispectral images. *Remote Sensing*, 14(22):5820.

Kaza, S., Yao, L., Bhada-Tata, P., and Van Woerden, F. (2018). *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Urban Development. World Bank Publications.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kumsetty, N. V., Nekkare, A. B., Sowmya Kamath, S., and Anand Kumar, M. (2023). An approach for waste classification using data augmentation and transfer learning models. In Kumar Singh, K., Bajpai, M. K., and Sheikh Akbari, A., editors, *Machine Vision and Augmented Intelligence*, pages 357–368, Singapore. Springer Nature Singapore.

Levenshtein, V. I. et al. (1966). Binary codes capable of correcting deletions, insertions, and reversals. 10(8):707–710.

Li, P., Wang, X., Su, M., Zou, X., Duan, L., and Zhang, H. (2021). Characteristics of plastic pollution in the environment: a review. *Bulletin of environmental contamination and toxicology*, 107:577–584.

Liang, S. and Gu, Y. (2021). A deep convolutional neural network to simultaneously localize and recognize waste types in images. *Waste Management*, 126:247–257.

Liu, K.-H., Chen, G.-H., and Liu, T.-J. (2021). Mix attention based convolutional neural network for clothing brand logo recognition and classification. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3013–3018. IEEE.

Longo, E., Sahin, F. A., Redondi, A. E., Bolzan, P., Bianchini, M., and Maffei, S. (2021). Take the trash out... to the edge. Creating a Smart Waste Bin based on 5G Multi-access Edge Computing. pages 55–60.

Lorang, S., Yang, Z., Zhang, H., Lü, F., and He, P. (2022). Achievements and policy trends of extended producer responsibility for plastic packaging waste in europe. *Waste Disposal & Sustainable Energy*, 4(2):91–103.

Lprdosmil (2022). Unsplash random images collection. Accessed: 2024-05-12.

Lu, W. and Chen, J. (2022a). Computer vision for solid waste sorting: A critical review of academic research. *Waste Management*, 142:29–43. Publisher: Elsevier.

Lu, W. and Chen, J. (2022b). Computer vision for solid waste sorting: A critical review of academic research. *Waste Management*, 142:29–43.

Mahat, S., Yusoff, S. H., Zaini, S. A., Midi, N. S., and Mohamad, S. Y. (2018). Automatic Metal Waste Separator System In Malaysia. pages 366–371.

Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykow, M. A., Kwasigroch, A., and Majek, K. (2022). Deep learning-based waste detection in natural and urban environments. *Waste Management*, 138:274–284.

Müller, R., Kornblith, S., and Hinton, G. E. (2019). When does label smoothing help? *Advances in neural information processing systems*, 32.

Nash, J. and Bosso, C. (2013). Extended producer responsibility in the united states: Full speed ahead? *Journal of Industrial Ecology*, 17(2):175–185.

OECD (2016). *Extended Producer Responsibility.*

Oluwadipe, S., Garelick, H., McCarthy, S., and Purchase, D. (2021). A critical review of household recycling barriers in the united kingdom. *Waste Management & Research*, page 0734242X211060619.

Park, J., Díaz-Posada, N., and Mejía-Dugand, S. (2018). Challenges in implementing the extended producer responsibility in an emerging economy: The end-of-life tire management in colombia. *Journal of Cleaner Production*, 189:754–762.

Pouikli, K. (2020). Concretising the role of extended producer responsibility in european union waste law and policy through the lens of the circular economy. 20(4):491–508.

Rahman, M. W., Islam, R., Hasan, A., Bithi, N. I., Hasan, M. M., and Rahman, M. M. (2022). Intelligent waste management system using deep learning with iot. *Journal of King Saud University-Computer and Information Sciences*, 34(5):2072–2087.

Ramasubramanian, B., Tan, J., Chellappan, V., and Ramakrishna, S. (2023). Recent advances in extended producer responsibility initiatives for plastic waste management in germany and uk. *Materials Circular Economy*, 5(1):6.

Ramirez, I., Cuesta-Infante, A., Pantrigo, J. J., Montemayor, A. S., Moreno, J. L., Alonso, V., Anguita, G., and Palombarani, L. (2020). Convolutional neural networks for computer vision-based detection and recognition of dumpsters. *Neural Computing and Applications*, 32:13203–13211.

Sheng, T. J., Islam, M. S., Misran, N., Baharuddin, M. H., Arshad, H., Islam, M. R., Chowdhury, M. E., Rmili, H., and Islam, M. T. (2020). An internet of things based smart waste management system using lora and tensorflow deep learning model. *IEEE Access*, 8:148793–148811.

Shulgin, M. and Makarov, I. (2023). Scalable zero-shot logo recognition. *IEEE Access*.

Somlai, C., Bullock, C., and Gallagher, J. (2023). Plastic packaging waste in europe: Addressing methodological challenges in recording and reporting. *Waste Management & Research*, 41(6):1134–1143.

Stanton, T., Chico, G., Carr, E., Cook, S., Gomes, R. L., Heard, E., Law, A., Wilson, H. L., and Johnson, M. (2022). Planet patrolling: A citizen science brand audit of anthropogenic litter in the context of national legislation and international policy. *Journal of Hazardous Materials*, 436:129118.

Trappey, A. J., Trappey, C. V., and Lin, E. (2022). Intelligent trademark recognition and similarity analysis using a two-stage transfer learning approach. *Advanced Engineering Informatics*, 52:101567.

Tumu, K., Vorst, K., and Curtzwiler, G. (2023). Global plastic waste recycling and extended producer responsibility laws. *Journal of Environmental Management*, 348:119242.

Wagner, R. A. and Fischer, M. J. (1974). The string-to-string correction problem. *Journal of the ACM (JACM)*, 21(1):168–173.

Walls, M. (2006). Extended producer responsibility and product design: economic theory and selected case studies.

Wang, C., Qin, J., Qu, C., Ran, X., Liu, C., and Chen, B. (2021). A smart municipal waste management system based on deep-learning and internet of things. *Waste Management*, 135:20–29.

Watkins, E. D., Gionfra, S., Schweitzer, J.-P., Pantzar, M., Janssens, C., and ten Brink, P. (2017). Epr in the eu plastics strategy and the circular economy: A focus on plastic packaging.

Wen, S., Yuan, Y., and Chen, J. (2023). A vision detection scheme based on deep learning in a waste plastics sorting system. *Applied Sciences*, 13(7).

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1, pages 29–39. Manchester.

Zdenek, J. and Caicedo, D. (2021). Crnn.

Zhang, F.-L., Du, S.-H., Guo, Z., et al. (2013). Extracting municipal solid waste dumps based on high resolution images. *Spectroscopy and Spectral Analysis*, 33(8):2024–2030.

Zhu, J., Fan, C., Shi, H., and Shi, L. (2019). Efforts for a circular economy in china: A comprehensive review of policies. *Journal of industrial ecology*, 23(1):110–118.

Appendix A. KNN evaluation report on DST dataset

Table A.2: Report of model performance on DST dataset with 23 classes. Imgs. is the number of images for each class.

	Precision	**Recall**	**F1-score**	**Imgs.**
Equori	0.74	0.97	0.84	40
Tosh	0.96	0.84	0.90	63
Jet	0.89	0.63	0.74	49
Oka Loka	0.72	0.72	0.72	40
Piazza	0.82	0.90	0.86	40
Kattakao	0.82	0.80	0.81	45
Chocorramo	0.68	0.82	0.74	39
Detodito	0.93	0.72	0.81	39
Galletas Dux	0.19	0.49	0.27	41
La Especial	0.85	0.77	0.80	43
Alpina	0.80	1.00	0.89	43
Choclitos	0.80	0.82	0.81	39
Chokis	0.70	0.81	0.75	43
Club Social	0.76	0.52	0.62	42
Del Valle	0.98	0.80	0.88	65
Festival	0.52	0.36	0.43	44
Kryzpo	0.82	0.44	0.57	32
Minichips	0.59	0.55	0.57	31
Nuthos	0.95	0.83	0.88	64
Quaker	0.96	0.86	0.91	51
Speed Max	1.00	0.78	0.88	36
Tostacos	0.80	0.92	0.85	38
Trululu	0.78	0.68	0.73	41
Accuracy			0.75	1008
Macro avg.	0.78	0.74	0.75	1008
Weighted avg.	0.80	0.75	0.76	1008

Appendix B. Hyperparameter exploration results

Figure B.9 shows the results of the hyperparameter exploration. Each circle represents a model instance, and the color scale indicates the macro

F1-score evaluated on the DST dataset. Each model instance is the best of 25 tries with the same hyperparameters but with randomly selected synthetic samples.

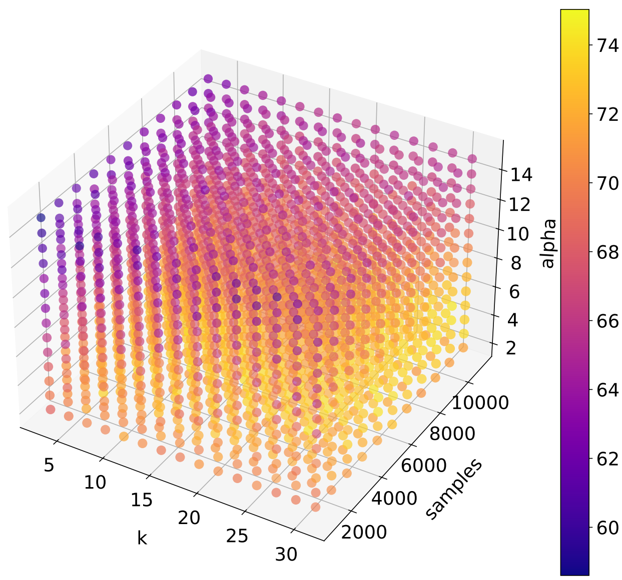

Figure B.9: Grid search exploration of α, number of copies, and k model hyperparameters.

Appendix C. Load test server's technical specifications

Table C.3: Server Load test specifications

Architecture	x86_64
Cores	8
vCPU	16
Sustained clock speed (GHz)	3.6
Threads per core	2
Memory (GiB)	32