Waste Management 119 (2021) 30–38

Contents lists available at ScienceDirect

Waste Management

journal homepage: www.elsevier.com/locate/wasman

Detecting glass and metal in consumer trash bags during waste collection using convolutional neural networks

Oliver Istad Funch, Robert Marhaug, Sampsa Kohtala *, Martin Steinert

Norwegian University of Science and Technology (NTNU), Department of Mechanical and Industrial Engineering, Richard Birkelands Veg 2B, 7491 Trondheim, Norway

ARTICLE INFO

Article history:
Received 30 June 2020
Revised 29 August 2020
Accepted 20 September 2020
Available online 8 October 2020

Keywords:
Trash bag classification
Convolutional neural network
Waste collection
Sound recognition
Metal detection

ABSTRACT

We present a proof-of-concept method to classify the presence of glass and metal in consumer trash bags. With the prevalent utilization of waste collection trucks in municipal solid waste management, the aim of this method is to help pinpoint the locations where waste sorting quality is below accepted standards, making it possible and more efficient to develop tailored procedures that can improve the waste sorting quality in areas with the most urgent needs. Using trash bags containing various amounts of glass and metal, in addition to common waste found in households, we use a combination of sound recording and a beat-frequency oscillation metal detector as inputs to a machine learning algorithm to identify the occurrence of glass and metal in trash bags. A custom-built test rig was developed to mimic a real waste collection truck, which was used to test different sensors and build the datasets. Convolutional neural networks were trained for the classification task, achieving accuracies of up to 98%. These promising results support this method's potential implementation in real waste collection trucks, enabling location-specific and long-term monitoring of consumer waste sorting quality, which can provide decision support for waste management systems, and research on consumer behavior.

1. Introduction and background

The increasing focus on preserving the global environment has led to marked changes in the municipal waste management policies around the world. The EU has set targets for the recycling of household waste at 50% by 2020, 55% by 2025, 60% by 2030, and 65% by 2035 according to Avfall Norge, the field association for recycling in Norway (Wilsgaard, 2018). Currently, this number is approximately 35% according to Renovasjonsetaten (REN), the department responsible for municipal waste management in Oslo.

According to Adhithya Prasanna et al. (2018), waste should be sorted at the earliest stage possible to reduce contamination during recycling processes. Improper sorting can lead to bag rupture due to sharp edges, contamination of plastic, which reduces the recyclability (J. Almankaas and J.H. Ellefsen-Killerud, personal communication, January 4, 2020), contaminated biogas production (Jiang et al., 2020), and contamination during waste incineration. Recycling materials, such as aluminum, is also a preferable alternative to incineration as the energy costs of aluminum production are much greater than the cost of recycling. The current municipal waste system in Oslo relies on home sorting by the consumer. Consumers sort their waste into bags of specific colors, which are later separated in a waste sorting facility. Neither glass nor metal should go in these bags as a separate collection point is dedicated to these materials. For this system to function, it is vital that the consumer correctly performs this sorting (Rousta and Ekström, 2013), as the sorting facility only considers the distinct bag color and not its content. Radio frequency identification (RFID) tags are fixed on the waste containers outside homes, which are read by the waste collection trucks to record the location of each disposal, along with a timestamp and the weight of each waste bin. The degree to which the contents of the colored bags are correctly sorted is not known during this collection process. At present, knowledge on incorrect sorting in Oslo is gathered by manually analyzing a portion of the collected waste once per year. In 2019, 4259 kg of waste from 10 different areas were analyzed and reported ("Avfallsanalysen 2019," 2019). This report found that, in the mixed waste category that will ultimately be incinerated, only 26.6%, in weight, of the contents belonged in that category. Of the materials not belonging in this category, the fourth largest was glass and metal, constituting 6.1% in weight, with the top three being food, paper, and plastic. There was also a significant difference in the sorting quality among the geographical areas, where the amount of incorrectly sorted trash varied from 9.8 to 63.4% in weight.

* Corresponding author.
E-mail addresses: sampsa.kohtala@ntnu.no (S. Kohtala), martin.steinert@ntnu.no (M. Steinert).

https://doi.org/10.1016/j.wasman.2020.09.032
0956-053X/

Nomenclature

BFO	Beat-Frequency Oscillation
CNN	Convolutional Neural Network
GM	Glass and Metal
GMX	Glass and Mixed waste
HMM	Hidden Markov Model
MFCC	Mel Frequency Cepstral Coefficients
ML	Machine Learning
MMX	Metal and Mixed waste
PG	Pure Glass
PM	Pure Metal
PMX	Pure Mixed waste
REN	Renovasjonsetaten
RFID	Radio Frequency Identification
ROC	Receiver Operating Characteristics
SVM	Support Vector Machine

Previous studies have attempted to improve upon consumer sorting, either by developing an automatic system to perform the sorting instead of the consumer or a system that places quality controls on home sorting tasks, which typically relies on certain type of classification methods. Convolutional Neural Networks (CNNs) have been heavily used in multiple waste classification applications after gaining popularity throughout the past few decades, mostly due to their impressive accuracy at classifying images with an increasing training speed (Krizhevsky et al., 2012). In one of the earlier studies that focused on applying machine learning to classify solid waste, Yang and Thung (2016) generated a dataset, known as TrashNet, with roughly 2400 images of solid waste in six different categories (glass, paper, metal, plastic, cardboard, and general trash). They achieved an accuracy of 63% using an SVM (support vector machine) with a feature detection algorithm, and 22% using a CNN. Bircanoğlu et al. (2018) experimented with several different CNN architectures on the TrashNet dataset. They achieved the highest accuracy of 95% by fine-tuning a pre-trained model and using data augmentation (flipping and rotating training samples) to increase the size of the dataset. Similarly, using randomly initiated weights for the model, Ruiz et al. (2019) achieved an accuracy of 89%. Lindermayr et al. (2018) extended the TrashNet dataset by capturing more images of trash. To resemble their application of roadside trash detection, they also generated synthetic data by segmenting the trash and adding different backgrounds to the images. Their best model reached an accuracy of 84% for the more challenging dataset. Toğaçar et al. (2020) combined an auto encoder, CNNs, and an SVM to classify images of waste as organic or recyclable from a large dataset of over 20,000 images, achieving nearly 100% accuracy. Chu et al. (2018) developed a classification system utilizing images together with weight and metal detection sensors. Their method, known as the multilayer hybrid method, used a CNN to extract image features and sensors to capture numerical features. Their model had an accuracy of over 90% when classifying items such as paper, plastic, metal, glass, and food waste as recyclable or not. Nowakowski and Pamuła (2020) proposed a classification model for electronic waste, including the ability to estimate the object size (object detection), yielding an accuracy of over 90%. Korucu et al. (2016) used a Hidden Markov Model (HMM) and SVM to classify materials based on sound. Their approach is similar to ours, although their focus lies in the source separation of packaging waste in reverse vending machines. They achieved up to a 100% classification accuracy when measuring free falling impact sounds from glass, plastic, metal, and cardboard, in addition to a high accuracy when estimating the size of the objects. Gong et al. (2019) demonstrated the use of microphone, accelerometer, and gyroscope data generated from a smartphone to recognize objects using an SVM. By simply knocking the phone against an object, they were able to identify the material with high accuracy, mainly due to the sound data.

The most prevalent shortcoming of the waste classification methods proposed in previous studies is that they are based on single objects per datapoint (Adhithya Prasanna et al., 2018), thus assuming that each individual piece of trash is separated beforehand, in addition to often lacking a description of direct real-world applications. Meanwhile, convenience is one of the driving factors for source-segregation of waste in households (Bernstad, 2014; Rousta et al., 2017), such as having separate bins for different waste types. While it is valuable to develop methods to automatically classify any type of waste, it is also important to consider their applicability in today's waste management systems to achieve the rapidly approaching goals set by the EU.

The selected approach in this study aims to identify glass and metal in collected waste without the need to manually inspect or separate the contents of individual bags. Glass and metal were chosen as they are commonly sorted incorrectly by consumers who are not using (or are not aware of) the designated collection points for these materials, as well as because these materials should never be found in the waste collected by the trucks. In addition, glass and metal pose a high risk of causing bag ruptures during the collection procedure, potentially increasing the level of contamination. Our method utilizes a combination of sound recording and a beat-frequency oscillation metal detector as inputs to a machine learning algorithm to identify the occurrence of glass and metal in trash bags. The intended implementation of this method is in trash collection trucks, with the aim of collecting information on the quality of sorting during normal collection routines. As a system with RFID is already in place for the trash bins scheduled for collection, our approach aims to help pinpoint the locations where the waste sorting quality is below accepted standards, making it possible and more efficient to develop tailored procedures that can improve the waste sorting quality in areas of most urgent need, thus improving the overall waste management system. We also discuss the potential benefits associated with collecting this type of data.

We develop an experimental design, including a custom-built test rig to capture datasets of trash bags containing different amounts of glass and metal. For the proof-of-concept trash-classification system, we generated bags in six different categories to train and evaluate several machine learning (ML) models. Multiple sensors are tested through an ablation study to obtain the best combination of input data for the models. Based on the results of the ablation study, several ML-models were trained and the final models are presented with their results and a discussion on the implications and prospects of our approach for supporting waste management systems.

2. Theory and methods

2.1. CNN, sound, and metal detection

A CNN is a supervised learning method, mapping two-dimensional input data (e.g., images) to output data (classes or categories). The main idea behind CNNs is to automatically learn to extract relevant features and find patterns from the input data

(LeCun et al., 1998). In simple terms, a CNN consists of an input and output layer, with hidden units in-between, consisting of convolutional- and pooling layers to extract features, and a fully connected neural network to calculate class probabilities based on these features. Using the input and output examples (i.e., labeled images) the CNN is able to update its weights via backpropagation to improve the classification accuracy of the model being trained. One of the challenges with CNNs is acquiring the large amount of data that is often required to properly train and tune a model. To resolve this challenge, a custom test rig was developed to enable rapid data acquisition and accelerate the testing of our concept.

Several studies have shown that sound recognition is an appropriate method to determine the material properties of objects. Giordano and McAdams (2006) reported good results on material recognition by humans based on impact sounds between gross material categories, but found that the acoustic and source properties contain sufficient information to identify even sub-categories of the same materials. Consequently, a machine learning algorithm may be able to perform better than a human by detecting small variations in the data. CNN-based algorithms typically employ spectrograms to process and visualize sound data and to highlight distinguishing features. Two widely used types are the Mel spectrogram and Mel Frequency Cepstral Coefficients (MFCC) spectrogram. To create a Mel spectrogram, a Fourier transform is applied to time segments of the sound clip, which shows the energy present for each frequency. The frequency axis is then scaled to a Mel scale, which is a log scale created to mimic how a human experiences sound (Volkmann et al., 1937). The energy axis is also scaled to a decibel scale, as the experienced sound volume is not linear (Chapman, 2000). The MFCC spectrogram is similar to the Mel spectrogram, but includes one additional processing step using a reverse Fourier transform, which results in a Cepstrum. This spectrogram has peak values where there are periodic elements in the time segment (Noll, 1967). Traditional sound recognition often utilizes an HMM, gaussian mixture model, or SVMs (Ananthi and Dhanalakshmi, 2015; Deng and Yu, 2014; Li et al., 2017; McLoughlin et al., 2015; Mesaros et al., 2010; Sharan and Moir, 2016), but several studies have also shown promising results using CNNs (Hershey et al., 2017; Khamparia et al., 2019; Kumar and Raj, 2017). CNN models often perform better in chaotic environments (Zhang et al., 2015) and are able to extract more abstract features while being unaffected by local variations (Çakır et al., 2017). In the case of trash collection, a significant amount of irregular acoustic noise can be expected, for which a CNN model may perform well.

Metal detection is a well proven concept widely used in multiple applications. Common methods include Beat-Frequency Oscillation (BFO), very low frequency, and pulse induction. In this study we used a BFO detector, mainly due to its simple design. The BFO creates an oscillating frequency using an LC-circuit. The LC-circuit causes an oscillation frequency due to the capacitor first discharging to the inductor, thereafter being charged by the induced voltage created in the coil. The charging and discharging are time-shifted between the components, which causes the current to rise and fall. As the internal resistance of the components cause the energy to dissipate, the circuit requires an external power supply to maintain oscillation. The frequency depends on the inductance of the search coil of the detector, which is obtained by winding the insulated cable in a loop. If a metal object is in close proximity to the coil, the inductance changes, which, in turn, changes the oscillation frequency. By storing this frequency and constantly comparing it to the current frequency, any change will indicate the presence of metal. An Arduino microcontroller can be used to perform this comparison and output the detection results.

3. Experimental setup and data recording procedures

A custom test rig was built to record sound and metal detector data from trash bags. The purpose of this rig is to mimic the area of the truck where the trash bins are emptied during the normal collection cycle, rendering data acquisition similar to real conditions. Fig. 1 shows a comparison of the tray in a waste collection truck and the landing tray of the test rig. Apart from the difference in size, the trays are similar in appearance. This simplified approximation provides increased control over the experiments in addition to enabling rapid prototyping of the classification system.

The main parts of the rig consist of a loading tray, chute, and steel landing tray. The chute is approximately the same size as a normal trash bin to simulate the velocity of trash bags before impact. A distance sensor is located at the top of the chute to register the passage of a trash bag. A Zoom H6n recorder is mounted next to the landing tray, which functions as a sound board, including two stereo condenser microphones attached to the recorder and two additional contact microphones placed on each side of the landing tray. A length of wire is wound around the chute, which acts as the search coil for the metal detector. The setup also includes a weight sensor beneath the loading tray, a GoPro camera mounted at the end of the landing tray, and a circuit board, containing an Arduino Nano and a HX711 load cell amplifier. Fig. 2 illustrates the rig.

Trash bags were created based on six different categories: pure metal (PM), pure glass (PG), metal and mixed waste (MMX), glass and mixed waste (GMX), glass and metal (GM), and pure mixed waste (PMX). Each category contains ~500 measurements to ensure sufficient data was obtained prior to analysis. As glass and metal are generally found mixed with other waste in a real-world scenario, PM and PG were omitted from the dataset, reducing it to ~2000 samples. Care was taken to ensure that all bags had different sizes, weights, and compositions within each category to simulate real conditions. The glass used was mostly bottles and jars that are normally found in households. The metal is an assortment of beverage cans and tins from canned food and some scrap sheet metal. The mixed waste category was composed of waste found in trash bins around the campus of the Norwegian University of Science and Technology, where the experiment was conducted, including different variations of empty food containers, such as cardboard, plastics, paper cups, and candy wrappers, and food waste, such as banana peels and apple cores. These materials were thoroughly inspected to ensure that no metal or glass were present. The MMX and GMX categories were created to ensure the inclusion of samples with high (50–70%), medium (20–50%), and low (5–20%) metal or glass content relative to the mixed waste to cover numerous possible scenarios and obtain more of the variance in the sorting behavior. Due to the extensive time requirements associated with producing unique bags for the entire training set, each bag was reused several times. In total, there were ~20 unique bags for each category. We hypothesize that, when the same bag is used several times with random orientations, the resulting sound and metal detection characteristics will be different. To verify our hypothesis, a separate dataset (test set) was created containing 40 unique bags (10 for each category), where the bags were not reused.

Each bag was recorded individually when capturing the datasets. The weight was automatically recorded after the bag was placed in the loading tray. After tilting the loading tray to initiate the recording procedure, the sound and metal detector data were recorded automatically for 2 s after the distance sensor was triggered. Video recording started when the weight was measured, and later automatically edited to include only 2 s after the distance sensor was triggered. The procedure was monitored from a custom

Fig. 1. The landing tray of a waste collection truck during operation on the left and the test rig on the right.

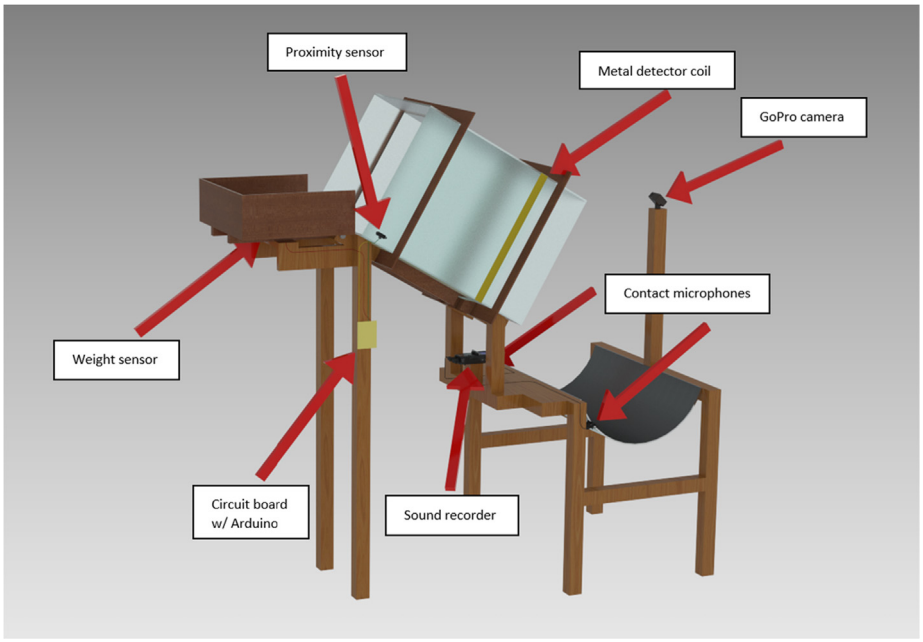

Fig. 2. A 3-D -model of the test rig showing all the components used for data collection.

graphical interface, allowing erroneous measurements to be discarded directly, rendering data collection more robust. More details on the custom test rig and data recording procedure can be found in the theses of Marhaug and Funch, n.d..

4. Data and model preparation

In this study, several CNN-based models were developed and tested. A simple ablation study was conducted to determine which sensors and input data contribute to model performance, including different sound recorders, sound data representations (Mel spectrogram and MFCC spectrogram), metal detector data, and weight. The ablation study was performed in consecutive order, where inputs contributing to a significant increase in accuracy (P-value below 0.05) were included in the next test. Consequent tests were compared to the previous best results.

Mel spectrograms and MFCCs were created using the Librosa package in Python. Examples of both spectrograms, including metal detection data, are shown in Fig. 3, each taken from the same sample. To reduce the sample size and computational power required for training, an algorithm was used to detect the moment of impact to extract a smaller frame around the event. The time window of 2 s was originally divided into 1379 columns, or time segments. Sixty-four segments before the impact and 192 segments after the impact were extracted for a total of 256 columns for each sample in the datasets, as we assumed that more features of interest would occur following impact. The vertical axis was resized to 256 using bilinear interpolation, resulting in a shape of 256×256. The metal detector data originally recorded 600 data points in a 1-D -array, totaling 2 s of data. This is longer than required as the bags only briefly pass the metal detector during the initial part of the recording; however, we preferred to ensure that all bags would be captured regardless of the sliding speed through the chute. The data was interpolated to 256 data points and repeated along the second axis, as the CNN requires a constant input shape for all channels. This approach was selected during training, as we observed that zero padding for the metal detection data resulted in the model effectively ignoring this input. The dataset was split into 80% for training and 20% for validation, totaling 1616 training samples and 403 validation samples, in addition to the test set containing 40 samples. All data from each measurement was input to the network as separate channels, where the

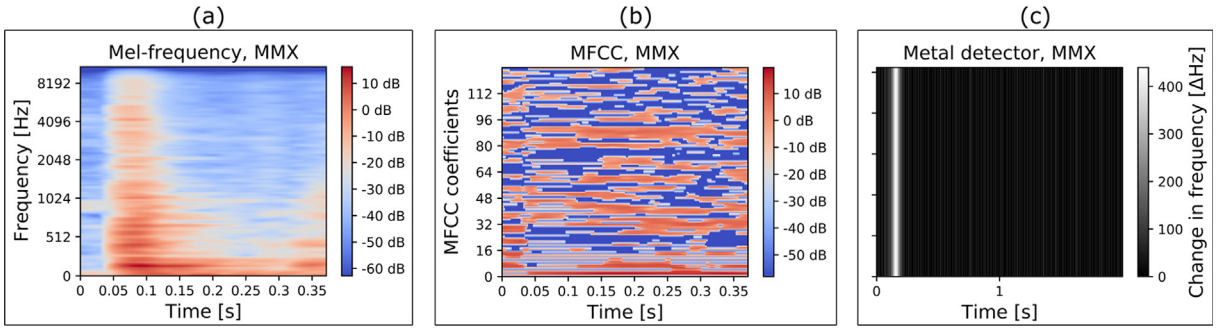

Fig. 3. Examples of a (a) Mel spectrogram, (b) MFCC, and (c) the transformed metal detector data for a sample from the metal and mixed waste category, used as input for the CNN.

data presented in Fig. 3 results in nine channels (two condenser- and contact microphones for each spectrogram, as well as the metal detector data).

Three CNN models with different numbers of layers and model parameters were tested. Model 1 (M-1) was a simple model having five convolutional layers, one fully connected layer, and uniformly sized kernels (3 × 3), which is based on common CNN architectures (Krizhevsky et al., 2012; LeCun et al., 1998; Simonyan and Zisserman, 2014). To assess the depth of the CNN models, model 2 (M-2) extended M-1 with two additional convolutional and fully connected layers while also having descending kernel sizes. Model 3 (M-3) was between M-1 and M-2 with respect to the layer count, but used a rectangular kernel (4 × 8) in the first layer to test if obtaining more information along the time domain while being concentrated along the frequency band can improve the classification accuracy. M-1 was the subject for the ablation study. In addition, several convolutional layers were tested for M-1, where we found that five layers was optimal for this dataset. Three labelling schemes were considered (multi-class, multi-label, and binary classification), where multi-labeling was selected due to its ability to consider each material independently, regardless of their combination in each sample.

To evaluate and discuss the classification performance, we used standard metrics commonly used to evaluate ML-models, i.e., accuracy, precision, and recall. Receiver operating characteristic (ROC) curves were also included, which offers a simple representation of the classification thresholds and their effect on true positive and false positive rates. The influence of the number of training samples is presented in the form of learning curves for each dataset, which can be analyzed to detect the degree of bias and variance in the models.

5. Results

5.1. Ablation study with input data

Table 1 lists the results of the ablation study. Initially, every input data was included as a reference. Weight was excluded in the first test, resulting in improved model performance. The next test (no metal) was then compared to the model trained with all the data, except for the weight. Excluding the metal detection data did not yield a significant difference in the accuracy for the validation set, despite the fact that the accuracy was substantially lower for the test set. Due to this significant reduction in the accuracy for the test set, metal detection data were included in both cases during further experimentation. The use of only one of the two microphones was compared in the third test, resulting in a contradiction between the datasets. For the validation set, the use of only the condenser microphone yielded a significant reduction in the accuracy while the test set indicates an even more significant increase in the accuracy. The difference is arguably small for the validation set while an increase of 2.25% in the accuracy is beneficial for the test set. The use of only the condenser microphone was therefore deemed favorable for our model. A reduction in the accuracy when only using the Mel spectrograms or MFCC was observed for the validation set, with no significant increase in the accuracy when using only the Mel spectrogram for the test set, thus showing the benefit of including both.

6. Results for final sensors setup

M-1, M-2, and M-3 were trained using the input data resulting in the highest increase in the accuracy from the ablation study. For each model, training was performed 20 times. Small variations were observed between training iterations, where Table 2 presents the best results. All results are based on a 0.5 threshold (or confidence) for counting a prediction as valid. M-1 achieves the highest validation accuracy, which is able to correctly predict the presence of metal at 100% accuracy and 96.28% for glass. Inference is also included for M-1 on the test set, showing a slightly lower performance compared with the validation set. The results for the test data are listed at the bottom of Table 2.

Fig. 4 shows the confusion matrices for M-1 on the validation and test sets. Here, glass has the highest number of false positives while metal is correctly predicted for every sample. Glass has more false negatives than false positives for the validation set, indicating a slight bias towards predicting no glass.

Table 1
. Results from the ablation study showing the change in model performance based on input data for both the validation (valid) and test sets. Every positive result is carried over to the next test, thus showing the cumulative changes in the score. Significant results are shown in bold.

Test	Input data	Number of runs	Average score change (valid)	P (T ≤ t)	Average score change (test)	P (T ≤ t)
0	All data	20	0.00%	Reference	0.00%	Reference
1	No weight	20	**+0.70%**	0.0096	**+2.19%**	0.0294
2	No metal	20	−0.05%	0.3449	**−16.06%**	< 0.0001
3	Condenser microphone	20	**−0.63%**	0.0031	**+2.25%**	0.0025
	Contact microphone	20	−0.42%	0.0550	−2.00%	0.0105
4	Mel	20	**−1.84%**	< 0.0001	+0.50%	0.4794
	MFCC	20	**−2.54%**	< 0.0001	**−17.00%**	< 0.0001

Table 2
. Accuracy (A), precision (P), and recall (R) for each model on the validation set, with the same metrics for the best performing model on the test set.

		All			Metal			Glass		
Model	Dataset	A	P	R	A	P	R	A	P	R
M-1	Valid	**98.14**	99.49	96.77	100.00	100.00	100.00	96.28	98.95	93.56
M-2		97.52	97.28	97.77	99.75	100.00	99.50	95.29	94.63	96.04
M-3		97.02	96.56	97.52	99.75	100.00	99.50	94.29	93.24	95.54
M-1	Test	96.25	95.24	97.50	100.00	100.00	100.00	92.50	90.48	95.00

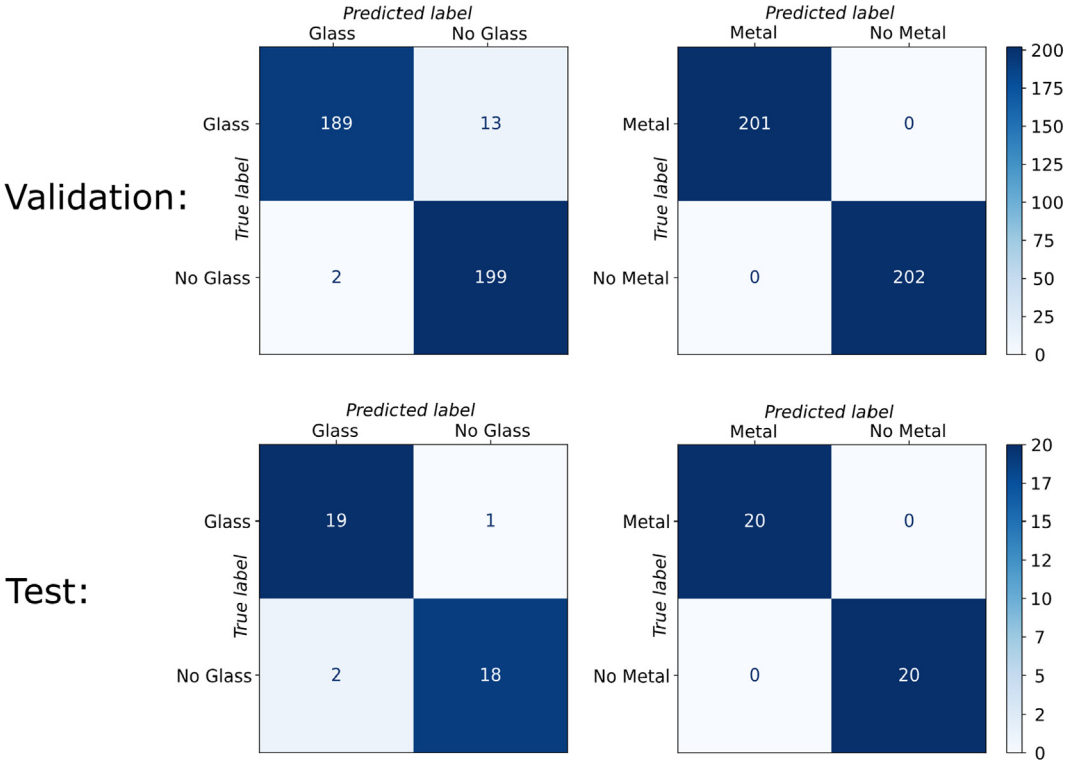

Fig. 4. Confusion matrices for glass and metal on each dataset. The columns show the number of respective labels predicted by the model while each row shows the correct label, constituting the four possible categories: true positive, false negative, false positive, and true negative. The color scale indicates the amount of predictions in each category.

Fig. 5 shows the receiver operating characteristics for M-1 for both the validation- and test sets. The AUCs near 100% show that there is little trade-off between the true positive- and false positive rate when tuning the threshold, such that a high recall can be achieved without sacrificing a significant amount of precision.

The learning curves in Fig. 6 appear to converge after including more than eight samples per category when training the model.

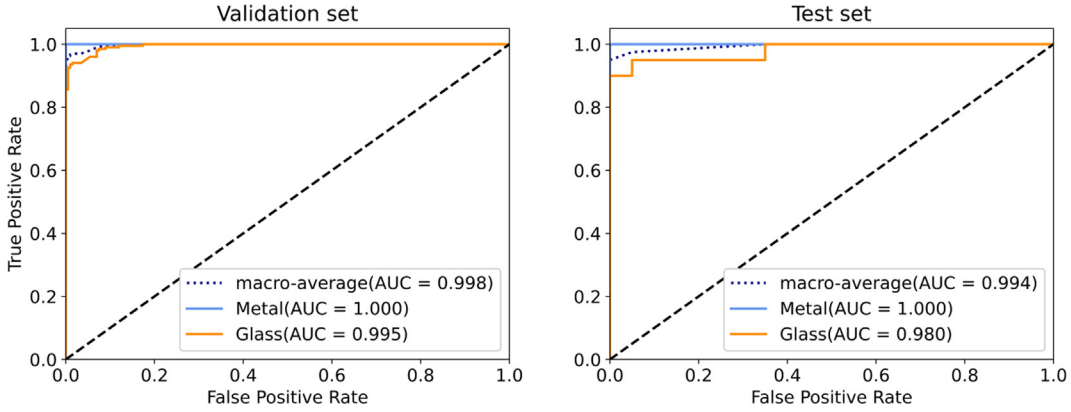

Fig. 5. ROC curves for the validation and test sets, including metal (blue curve), glass (orange curve), and the macro-average of both (dotted line), generated from testing the model with every possible threshold and reporting the corresponding true positive and false positive rates. The respective AUCs are indicated in the legend. The diagonal dashed line represents the performance expected of a random model, to clearly distinguish a well performing model (above this line).

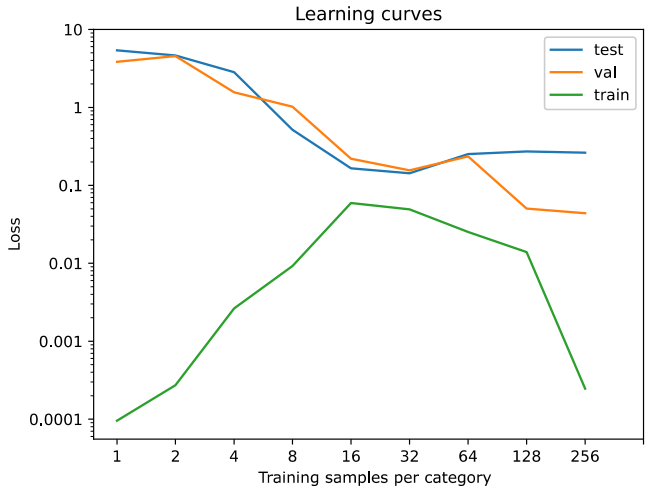

Fig. 6. Learning curves where the loss for each dataset is calculated for each increment of the number of training samples used. Both the loss-values (y-axis) and number of training samples per category (x-axis) use logarithmic scales. The learning curves for the test, validation, and training datasets are indicated by the blue, orange, and green lines, respectively.

With each dataset characterized by small loss values and convergence, few samples were required to train the model while showing low levels of bias and variance. Consequently, adding more training data does not improve the model, with optimal model performance achieved using only 16 samples per category (64 samples in total).

7. Discussion

Our results show significant potential for applying CNNs to classify the presence of glass and metal in trash bags using sound and metal detection, with each of the models achieving high accuracies. While Korucu et al. (2016) demonstrated the applicability of using sound to detect materials in a reverse vending machine, our approach takes this application a step further by detecting mixed materials inside the trash bags. The accuracy of our models (96.25–98.14%) is also comparable to Korucu et al. (2016), who achieved 96.5–100% accuracy, despite the fact that scenarios and input data are different. Compared with the other methods used to classify individual waste objects (Bircanoğlu et al., 2018; Chu et al., 2018; Lindermayr et al., 2018; Nowakowski and Pamuła, 2020; Ruiz et al., 2019; Toğaçar et al., 2020; Yang and Thung, 2016), with accuracies ranging from 63 to 99.95%, we are well within the expected classification performances possible at present. Based on our objective of supporting current waste management systems that rely on consumer sorting, without proposing radical changes to the status quo, our approach appears to be feasible based on our results, where the process itself is more applicable as compared with previous studies (Bircanoğlu et al., 2018; Chu et al., 2018; Korucu et al., 2016; Lindermayr et al., 2018; Nowakowski and Pamuła, 2020; Ruiz et al., 2019; Toğaçar et al., 2020; Yang and Thung, 2016).

The ablation study consolidated the benefit of using both the Mel and MFCC spectrograms. We also found that using only condenser microphones, and not contact microphones, can improve detection, which is desirable as less hardware is required, thus reducing costs in potential future implementations. However, given that this is a proof-of-concept, we still recommend testing both microphones in a realistic setting as our dataset can be biased from the controlled experiment. A commercial metal detector should also be tested in an actual scenario, where a longer range and improved robustness may be required. Nonetheless, the use of metal detection data had less effect when running the models on the validation set, but had a significant contribution to improvements in the accuracy of the test set. Although metal was detected with 100% accuracy, making it possible to detect metal using only a simple threshold-based algorithm, it may be an important input for ML models when detecting finer details, such as metal lids on glass jars, which were not accounted for in our dataset. ML models may also be able to differentiate between low and high metal or glass contents in bags, providing more information on consumer sorting behavior.

Altering the model architecture appears to have little effect on the performance, where the baseline model (M-1) is slightly better. This model shows slightly better accuracy for the validation set, which contains bags from the same distribution as the training data where bags were reused, as compared to the test set containing only unique bags. Reusing bags with random orientation may have less variation than expected, causing the models to overfit for the bags used for training and validation. The test set may contain slightly different features due to data collection during a different time and location, which may have affected the sensor readings. Analyzing the variance between the sensor readings of the trash bags when using different orientations and different environments may be valuable to better understand if either approach can or should be used in the future. The benefit of reusing bags is to obtain more data faster because creating unique bags is a cumbersome and time-consuming task. However, based on the learning curves shown in Fig. 6, relatively few training samples may be needed, and capturing samples with large variation is more important. When approaching a real-world application, we suggest collecting trash bags from consumers to build the datasets. It is then, for example, possible to pass the collected bags through the experimental rig or a collection truck before manually analyzing the bags to determine the correct label for the data, thereby producing an even more realistic dataset for training a classifier. To reduce the manual labor, an image-based classification system can also be implemented for this purpose.

Adjusting the classification threshold typically results in either an increase in the precision and reduction in the recall, or vice versa. According to the ROC plots in Fig. 5, this trade-off is miniscule for our models; however, this is likely to be substantially larger (smaller AUC) for real-world applications. The selection of the metric depends on how the results will be used. If the goal is to initiate pecuniary measures based on those who sort poorly, then perhaps avoiding false positives is essential, thereby maximizing the precision. If the measures are implemented, for instance, to distribute more information in areas of poor sorting, then including as many of the cases as possible may be of more interest, thereby maximizing the recall.

An important factor to consider in a real-world situation is the ratio between the correctly and incorrectly sorted bags, which is ~20:1 according to statistics from REN (J. Almankaas and J.H. Ellefsen-Killerud, personal communication, January 4, 2020). Applying Bayes' theorem in the case of detecting glass using the results from the test set, we attempt to calculate the likelihood of finding a true positive (A) given that the model has classified the bag as positive (B), P(A|B). Here, the probability of classifying a positive given a true positive, P(B|A), is the precision (90.48%) while the probability of finding a true positive P(A) is 1/21, with a false positive rate of 0.1. Therefore, we obtain a 31.15% likelihood that a classified positive is a true positive. In other words, we can expect 2.2 false positives for every true positive. If multiple areas of trash collection over time are considered, this is an acceptable result as the positive detections are likely to accumulate at the locations where incorrect sorting occurs more often, thereby increasing the likelihood of locating the true positives.

The datasets used in this study were captured under highly controlled circumstances using one bag at a time. Given that this is a proof-of-concept using a simplified approach, it is unclear how these models will perform on data from the actual collection routine. One possible problem relating to the collection truck is that many bags are emptied at once, which may increase the classification difficulty. If this proves a problem, a possible solution may be to include a type of funnel on the collection truck, which forces the bags to fall through one at a time, or to simply train the model with multiple bags at a time. In either case, we recommend obtaining the training data from a real collection truck, which will contain more background noise that is difficult to account for with a custom setup. Detecting the presence of glass and metal in each individual bag is not necessarily crucial, as only the trash bins have known geographical locations and not the individual bags. For every trash bin emptied by the collection truck, if glass or metal is detected in at least one of the bags, it is a useful datapoint to pinpoint the general area where incorrect sorting occurs over time.

While this method has been mainly developed for use in a collection truck, there are other areas where its implementation would be easy and effective. For example, this method can support the sorting facility when separating bags based on their content, although information on the source of incorrect sorting behavior will be lost during this stage. In addition, numerous condominium buildings have shared trash disposal units, often hidden underground, where the chute provides an excellent location for the installation of sensors. These systems often have an RFID in place, enabling the establishment of a link between the consumer and their sorting behavior.

Obtaining data on consumer sorting behaviors is beneficial to most waste management systems as it enables fact-based decision making. Most households in Oslo have a distance of <300 m to their nearest collection point, although in certain cases it is further. Together with our approach, it is possible to optimize the location of collection points as a benefit to consumers because convenience is an important factor (Bernstad, 2014; Rousta et al., 2017). Certain waste management systems also utilize mobile recycling stations that accept, among other recyclables, metal and e-waste, which are regularly moved. Data on consumer sorting may aid in the selection of optimal routines and placements for these mobile units.

If successfully deployed, our approach may also support research on consumer waste sorting behaviors. In general, data obtained with our system may contribute to studies attempting to predict municipal solid waste generation (Adamović et al., 2018; Kannangara et al. (2018); Wu et al., 2020), as it may provide accurate location specific data. Rousta and Ekström (2013) examined the environmental, economic, and social aspects of incorrect sorting, concluding that future research should be conducted to uncover the driving factors of consumer sorting, for which our method may be useful. Bernstad (2014) argued that research on factors influencing participation in waste sorting has been largely inconsistent, where the effects of promotional campaigns are often unknown due to a lack of proper monitoring. Our approach enables long-term monitoring, which can be used to reveal the most sustainable methods to improve consumer sorting.

8. Conclusions

With the increasing focus on preserving the global environment, necessary targets have been set to increase the recycling of household waste within the next few years. Consumer sorting is widely deployed as the first step in a sustainable waste management system. As the degree of correct sorting is not known during the normal collection process, we aim to support waste management systems in understanding consumer waste generation and behavior by enabling the classification of trash bags during this process. This proof-of-concept system is able to identify the occurrence of glass and metal in consumer trash bags with high accuracy. With an RFID system in place for the collection of municipal waste, our method can help pinpoint the areas where incorrect sorting occurs without implementing considerable changes to the current system.

We were able to successfully develop a classification system that comprises a combination of sound recording and a beat-frequency oscillation metal detector. The trained CNN model can identify the occurrence of glass and metal in trash bags with an accuracy of 98%. Considering the experimental nature of this study, along with the high accuracies achieved with relatively small data requirements, the potential for the application of this method in actual situations is promising.

For future research, we suggest collecting more realistic datasets of consumer trash bags to train the CNN models. The use of sound recorders and metal detection has shown promising results and should be tested in more realistic settings. Enabling long-term monitoring of consumer waste sorting quality will also benefit research on consumer behavior. For waste management systems, fact-based decision making can be realized using our approach to, for example, optimize the locations of collection points to increase the convenience for consumers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Renovasjonsetaten in Oslo for providing valuable insights to their waste management system.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wasman.2020.09.032.

References

Adamović, V.M., Antanasijević, D.Z., Ćosović, A.R., Ristić, M.Đ., Pocajt, V.V., 2018. An artificial neural network approach for the estimation of the primary production of energy from municipal solid waste and its application to the Balkan countries. Waste Manage. 78, 955–968. https://doi.org/10.1016/j.wasman.2018.07.012.

Adhithya Prasanna, M., Vikash Kaushal, S., Mahalakshimi, P., 2018. Survey on identification and classification of waste for efficient disposal and recycling. Int. J. Eng. Technol. 7 (2.8), 520–523. https://doi.org/10.14419/ijet.v7i2.8.10513.

Ananthi, S., Dhanalakshmi, P., 2015. SVM and HMM modeling techniques for speech recognition using LPCC and MFCC features. In: Paper presented at the Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014.

Avfallsanalysen 2019, 2019. Retrieved May 23, 2020, from https://www.oslo.kommune.no/avfall-og-gjenvinning/hvordan-kildesortere-i-oslo/avfallsanalysen/#gref.

Bernstad, A., 2014. Household food waste separation behavior and the importance of convenience. Waste Manage. 34 (7), 1317–1323.

Bircanoğlu, C., Atay, M., Beşer, F., Genç, Ö., Kızrak, M.A., 2018. RecycleNet: Intelligent waste sorting using deep neural networks. Paper presented at the 2018 Innovations in Intelligent Systems and Applications (INISTA).

Çakır, E., Parascandolo, G., Heittola, T., Huttunen, H., Virtanen, T., 2017. Convolutional recurrent neural networks for polyphonic sound event detection. IEEE/ACM Trans. Audio, Speech, Language Process. 25 (6), 1291–1303. https://doi.org/10.1109/TASLP.2017.2690575.

Chapman, D.M.F., 2000. Decibels, SI units, and standards. J. Acoust. Soc. Am. 108 (2), 480. https://doi.org/10.1121/1.429620.

Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S., Xiong, X., 2018. Multilayer hybrid deep-learning method for waste classification and recycling. Comput. Intelligence Neurosci.

Deng, L., Yu, D., 2014. Deep Learning: Methods and Applications. Found. Trends Signal Process. 7 (3–4), 197–387. https://doi.org/10.1561/2000000039.
Giordano, B.L., McAdams, S., 2006. Material identification of real impact sounds: effects of size variation in steel, glass, wood, and plexiglass plates. J. Acoust. Soc. Am. 119 (2), 1171. https://doi.org/10.1121/1.2149839.
Gong, T., Cho, H., Lee, B., Lee, S.-J., 2019. Knocker: vibroacoustic-based object recognition with smartphones. Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol. 3 (3), 1–21. https://doi.org/10.1145/3351240.
Hershey, S., Chaudhuri, S., Ellis, D.P.W., Gemmeke, J.F., Jansen, A., Moore, R.C., Plakal, M., Platt, D., Saurous, R.A., Seybold, B., Slaney, M., Weiss, R.J., Wilson, K., 2017, March 2017. CNN architectures for large-scale audio classification. In: Paper presented at the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
Jiang, P., Fan, Y.V., Zhou, J., Zheng, M., Liu, X., Klemeš, J.J., 2020. Data-driven analytical framework for waste-dumping behaviour analysis to facilitate policy regulations. Waste Manage. 103, 285–295. https://doi.org/10.1016/j.wasman.2019.12.041.
Kannangara, M., Dua, R., Ahmadi, L., Bensebaa, F., 2018. Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. Waste Manage. 74, 3–15. https://doi.org/10.1016/j.wasman.2017.11.057.
Khamparia, A., Gupta, D., Nhu, N., Khanna, A., Pandey, B., Tiwari, P., 2019. Sound classification using convolutional neural network and tensor deep stacking network. IEEE Access PP. https://doi.org/10.1109/ACCESS.2018.2888882.
Korucu, M.K., Kaplan, Ö., Büyük, O., Güllü, M.K., 2016. An investigation of the usability of sound recognition for source separation of packaging wastes in reverse vending machines. Waste Manage. 56, 46–52.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Paper presented at the Advances in Neural Information Processing Systems.
Kumar, A., Raj, B., 2017. Deep CNN Framework for Audio Event Recognition using Weakly Labeled Web Data. arXiv:1707.02530 [cs].
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.
Li, J., Dai, W., Metze, F., Qu, S., Das, S., 2017, March 2017. A comparison of Deep Learning methods for environmental sound detection. In: Paper presented at the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
Lindermayr, J., Senst, C., Hoang, M.-H., Haegele, M., 2018. Visual classification of single waste items in roadside application scenarios for waste separation. In: Paper presented at the 50th International Symposium on Robotics (ISR 2018).
Marhaug, R., Funch, O.I., n.d.. *Detecting Improperly Sorted Materials in Trash Bags* (Master's thesis). Norwegian University of Science and Technology, Trondheim, Norway.
McLoughlin, I., Zhang, H., Xie, Z., Song, Y., Xiao, W., 2015. Robust sound event classification using deep neural networks. IEEE/ACM Trans. Audio, Speech, Language Process. 23 (3), 540–552. https://doi.org/10.1109/TASLP.2015.2389618.
Mesaros, A., Heittola, T., Eronen, A., Virtanen, T., 2010, August 2010. Acoustic event detection in real life recordings. In: Paper presented at the 2010 18th European Signal Processing Conference.
Noll, A.M., 1967. Cepstrum pitch determination. J. Acoust. Soc. Am. 41 (2), 293–309. https://doi.org/10.1121/1.1910339.
Nowakowski, P., Pamuła, T., 2020. Application of deep learning object classifier to improve e-waste collection planning. Waste Manage. 109, 1–9.
Rousta, K., Ekström, K.M., 2013. Assessing incorrect household waste sorting in a medium-sized Swedish city. Sustainability 5 (10), 4349–4361.
Rousta, K., Ordoñez, I., Bolton, K., Dahlén, L., 2017. Support for designing waste sorting systems: a mini review. Waste Manage. Res. 35 (11), 1099–1111.
Ruiz, V., Sánchez, Á., Vélez, J.F., Raducanu, B., 2019. Automatic image-based waste classification. Paper presented at the International Work-Conference on the Interplay Between Natural and Artificial Computation.
Sharan, R.V., Moir, T.J., 2016. An overview of applications and advancements in automatic sound recognition. Neurocomputing 200, 22–34. https://doi.org/10.1016/j.neucom.2016.03.020.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
Toğaçar, M., Ergen, B., Cömert, Z., 2020. Waste classification using AutoEncoder network with integrated feature selection method in convolutional neural network models. Measurement 153,. https://doi.org/10.1016/j.measurement.2019.107459 107459.
Volkmann, J., Stevens, S.S., Newman, E.B., 1937. A scale for the measurement of the psychological magnitude pitch. J. Acoust. Soc Am. 8 (3). https://doi.org/10.1121/1.1901999.
Wilsgaard, S., 2018. Europa har fått nye avfallsdirektiv. Retrieved May 12, 2019, from https://www.avfallnorge.no/bransjen/nyheter/europa-har-f%C3%A5tt-nye-avfallsdirektiv.
Wu, F., Niu, D., Dai, S., Wu, B., 2020. New insights into regional differences of the predictions of municipal solid waste generation rates using artificial neural networks. Waste Manage. 107, 182–190. https://doi.org/10.1016/j.wasman.2020.04.015.
Yang, M., & Thung, G. (2016). Classification of trash for recyclability status. CS229 Project Report, 2016.
Zhang, H., McLoughlin, I., Song, Y., 2015, April 2015. Robust sound event recognition using convolutional neural networks. In: Paper presented at the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).