LusTra: waste type prediction through sound and accelerometer data

Ludovic Pfeiffer
HEPIA HES-SO
University of Applied Sciences
Geneva, Switzerland
ludovic.pfeiffer@hesge.ch

Gabriel Da Silva Marques
HEPIA HES-SO
University of Applied Sciences
Geneva, Switzerland
gabriel.dasilva-marques@hesge.ch

Florent Glück
HEPIA HES-SO
University of Applied Sciences
Geneva, Switzerland
florent.gluck@hesge.ch

Noria Foukia
HEPIA HES-SO
University of Applied Sciences
Geneva, Switzerland
noria.foukia@hesge.ch

Maïck Huguenin-Vuillemin
HE-ARC HES-SO
University of Applied Sciences
Neuchâtel, Switzerland
maick.huguenin@he-arc.ch

Artan Sadiku
HE-ARC HES-SO
University of Applied Sciences
Neuchâtel, Switzerland
artan.sadiku@he-arc.ch

Nuria Pazos Escudero
HE-ARC HES-SO
University of Applied Sciences
Neuchâtel, Switzerland
nuria.pazosescudero@he-arc.ch

Olivier Hüsser
HE-ARC HES-SO
University of Applied Sciences
Neuchâtel, Switzerland
olivier.huesser@he-arc.ch

***Abstract*—In recent years, advancements in waste sorting have been significantly enhanced by the integration of deep learning algorithms. In this regard, LusTra proposes a waste type recognition system using sound and accelerometer data. Two waste types are considered: Polyethylene Terephthalate (PET) bottles and Aluminium cans.**

Predictions through sound data, converted to Mel spectrograms, with Convolutional Neural Networks (CNN), are promising and result in an accuracy of 89% for PET waste and 90% for aluminium waste.

Random forest with features extracted from the accelerometer data provide an accuracy of 76% and 86% for PET and aluminium respectively.

***Index Terms*—Machine learning, waste sorting, sound, accelerometer, convolutional neural network, random forest.**

I. Introduction

In recent years, the amount of incinerable waste in Switzerland has remained stable. However, recycling rates for recoverable materials have plateaued due to a lack of innovation and optimisation. While the current recycling efforts are helping to conserve resources and protect the environment, there is still room for improvement. Especially in reducing the amount of incorrectly sorted waste despite the availability of sorted (glass, paper, ...) recycling bins. Encouraging and teaching eco-friendly behavior is essential to address these sorting issues and must involve everyone, regardless of social status or age. This is where the LusTra (*Ludique*[1] *smart Trash*) project steps in.

LusTra is a solution that proposes a fun, motivating sorting system aimed at educating young people. Through an interactive and rewarding game, it aims to make recycling a natural habit, fostering better waste sorting practices and promoting environmental awareness among the younger generations.

The Lustra project has been funded with the support of the HES-SO. We would like to thank the HES-SO for its invaluable support.

[1]in French, means playful in english

For this purpose, LusTra combines three systems:

- A signal acquisition system.
- A machine learning (ML) classification system.
- A game feedback system.

One of LusTra objective is to adapt to any type of sorting bin. Since images acquisition requires an external lighting source, it greatly increases the complexity of a system that has to adapt to all types of bins. For this reason, a sound and an accelerometer signal acquisitions have been used instead of an images one.

After relating the existing work, this paper describes the signal acquisition system, the collected data, the theoretical results obtained by the classification system as well as the real-life ones.

II. Related work

Recently, advancements in waste sorting have been significantly enhanced by the integration of deep learning algorithms. Companies like AMP Robotics are leveraging modern AI techniques to quickly identify and recognize objects, often outperforming human capabilities [1].

Despite the increasing rates of recycling, the volume of waste generated continues to rise. [2] tackled this challenge and explored how AI-driven algorithms can enhance the efficiency of waste classification and separation. [3] conducted a critical review of academic research on computer vision (CV) enabled municipal solid waste sorting. They highlighted a shift from traditional ML to deep learning algorithms. In this respect, [4] developed an industrial machine that uses an automatic learning algorithm to recognize and sort different types of plastic as they move along a conveyor belt. Whereas [5] investigated the use convolutional neural networks (CNN) and support vector machines (SVM), to classify industrial waste into paper, plastic, and metal. [6] also conducted a study that utilized SVM and CNN to automate the sorting process between plastic, paper, and metal. They found that SVM outperformed CNN in accuracy, achieving 94.8% compared to 83%. [7] developed an autonomous waste sorting machine capable of detecting multiple classes of waste materials and separating them accordingly. This project demonstrates the potential of Region Based Convolutional Neural Networks (R-CNN) to significantly reduce human effort in waste sorting and make the process completely automatic. [8] proposed a hierarchical deep learning approach for detecting and classifying waste in food trays, introducing *Labeled Waste in the Wild* dataset, for research purposes. [9] proposed an automatic ML-based waste recycling framework to improve the separation of mixed recyclables. CleverTrash integrates a waste recognition system into recycle bins, providing sorting statistics to improve waste collection strategies[10].

Most of these previous works focus on waste image processing. However, [11] used a one-dimensional convolutional neural networks for waste sorting based on acoustic data with a classification accuracy of 92.40%. LusTra follows the same vein by determining the waste through sound and vibrations. Furthermore, LusTra offers a complete architecture, with data acquisition, processing and playful feedback to users.

III. Acquisition System

To generate data, a custom 3D-printed signal acquisition system (attached to the bin lid, see Fig.1) for sound, vibration and metal detection has been created. Sound and vibration metrics are measured from the impact of waste with a 45° plate 20cm under the bin entry. Acquisition for all sensors are triggered by a light barrier crossing on the bin lid entry. An accelerometer, placed underneath the plate, records Z accelerations and a microphone, next to the plate, records sounds. The metal detector, made with copper coils, is able to detect eddy currents on an aluminium (Alu) can. An embedded Raspberry pi (RPI) controls all sensors.

When the light barrier is crossed, the microphone, the accelerometer and the metal detector register data for one second. If no relevant acceleration is detected, data are discarded. If the metal detector identifies eddy currents, that means it is not PET and probably aluminium. The RPI executes the ML-based algorithms on sound and acceleration data and upload results via Wi-Fi to a dedicated server. Allowing us to produce statistics for the deployed prototype. Fig.2 shows LusTra overall system architecture.

IV. Data Acquisition

A. Data collection

The acquisition system needed for the ML-based classification collects two types of data. The sound data, captured by the microphone, as a WAV file, and the accelerometer data (only the Z-axis), captured by the accelerometer, as a binary file. The recording of these two signals is triggered by the light barrier and lasts one second.

As LusTra works with PET and aluminium sorting bins, it was necessary to determine the waste types to be collected. The decision was made to work with three different classes: PET waste; aluminium waste; other waste.

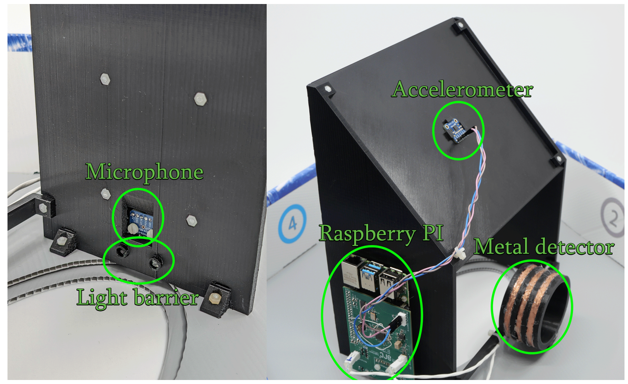


Fig. 1. LusTra full acquisition system

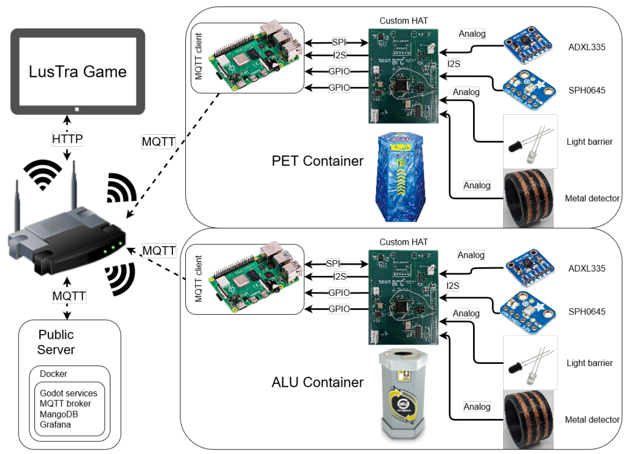


Fig. 2. LusTra overall system architecture

For each class, different shapes of waste of the same type are possible. For example, a PET bottle or an aluminium can may be intact, crushed, contain remnants of liquids, etc.

The remaining waste fall into the Other class. For instance: glass bottles, organic waste (banana peel, apple core), paper (crumpled, torn, intact), etc.

To build the dataset[2], 1,245 sound samples have been collected, distributed as follows: 379 aluminium cans; 295 Other; 571 PET bottles.

Considering that the system can potentially be in noisy places, some data have been recorded with environmental noise in addition to the impact on the system bar.

Changes made to the system during the project led to fewer accelerometer data. Therefore, the data acquired by the accelerometer at the beginning differ from the final data. The accelerometer data are divided as follows: 218 aluminium; 159 Other; 316 PET. Making a total of 693 samples.

B. Data pretreatment

Audio recordings in WAV format (displayed in Fig. 3) and accelerometer readings in binary format (Fig. 4) were

[2]Available here: https://doi.org/10/gt4kbk

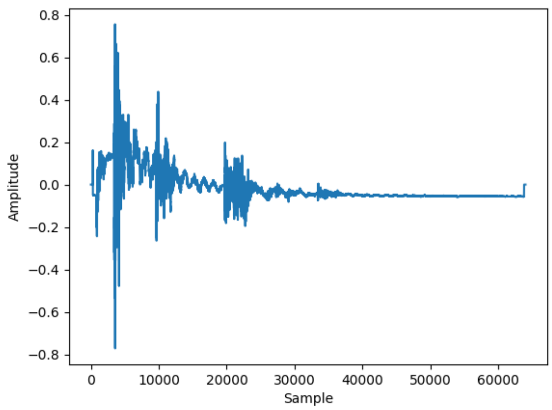


Fig. 3. Example of PET bottle sound

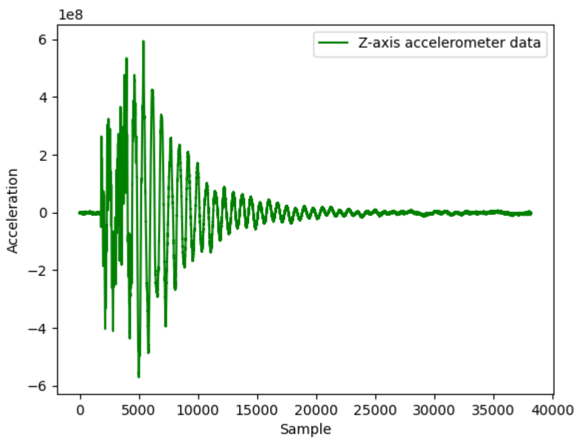


Fig. 4. PET accelerometer data example

collected. The next phase performs a comprehensive treatment of these datasets to extract meaningful insights.

After collection, the sound (64kHz) is converted into a Mel spectrogram, which visually represents the frequency spectrum on a Mel scale over time. This eases the identification of temporal and spectral patterns. Fig. 5 shows an example of a PET bottle Mel spectrogram.

For the accelerometer data (38400 Hz). The process is more straightforward because the accelerometer offset is removed (the offset value varies depending on the accelerometers, even if they are the same models). Fig. 4 shows an example of the accelerometer values after removing the offset.

V. Classifications and Results

Two different classifications have been defined: PET vs Other and Aluminium vs other. PET vs Aluminium vs Other has also been considered.

For each classification, the same seed has been used to control randomness and all data samples are normalized. Each training run uses the following distribution: Training set with

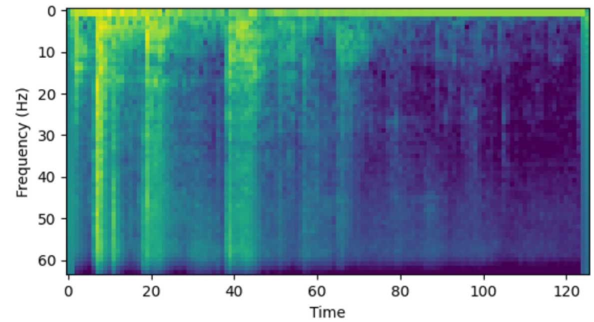


Fig. 5. Example of a PET Mel spectrogram

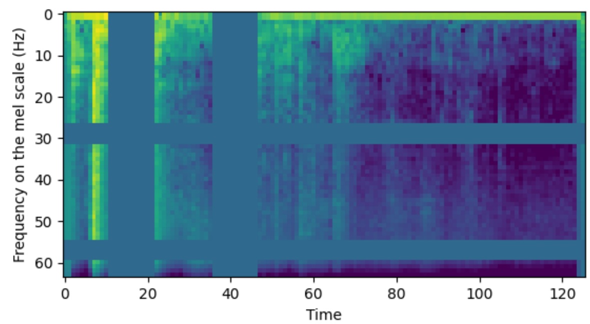


Fig. 6. PET Mel spectrogram with a masked section

80% of data (30% used for the validation set) and the testing set with the remaining 20%.

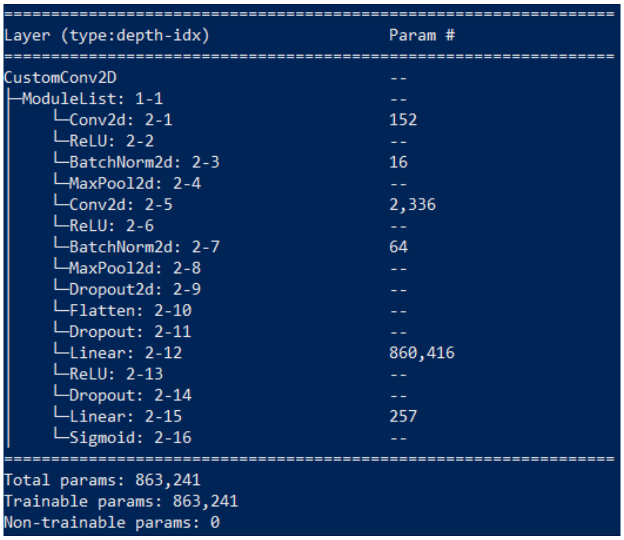
”Other” is considered the negative class (value 0) and ”PET” or ”Alu” the positive one (value 1).

A. Sound classification

As explained in section IV-B, all sounds are converted into a Mel spectrogram using the Pytorch function MelSpectrogram. The spectrogram can be influenced by three key parameters: the sound duration (up to 1s), the number of Fast Fourier Transforms (FFT), and the number of Mel filterbanks. Changes in these parameters can impact classification results.

With various combinations, the best results are obtained by keeping the duration at one second, the number of FFTs at 1024 and using 64 Mel filterbanks. To avoid overfitting, a random section of the spectrogram is removed (as shown in Fig. 6).

As the sound data are converted into a Mel spectrogram, a deep learning approach is used and those sounds are treated as images. Different models of varying depths and varying parameters have been tested before deciding on the best architecture. As with depth, the composition of the convolution blocks was tested: With different activation functions (ReLU, leakyReLU, hardswish...) and the inclusion or not of a batch normalization layer. Several optimisers have been used, including Adam and SGD, with different learning rates. To address overfitting, a varying number of dropout layers has



```
================================================================
Layer (type:depth-idx)                          Param #
================================================================
CustomConv2D                                    --
├─ModuleList: 1-1                               --
│    └─Conv2d: 2-1                              152
│    └─ReLU: 2-2                                --
│    └─BatchNorm2d: 2-3                         16
│    └─MaxPool2d: 2-4                           --
│    └─Conv2d: 2-5                              2,336
│    └─ReLU: 2-6                                --
│    └─BatchNorm2d: 2-7                         64
│    └─MaxPool2d: 2-8                           --
│    └─Dropout2d: 2-9                           --
│    └─Flatten: 2-10                            --
│    └─Dropout: 2-11                            --
│    └─Linear: 2-12                             860,416
│    └─ReLU: 2-13                               --
│    └─Dropout: 2-14                            --
│    └─Linear: 2-15                             257
│    └─Sigmoid: 2-16                            --
================================================================
Total params: 863,241
Trainable params: 863,241
Non-trainable params: 0
```

Fig. 7. Best PET/Alu model architecture

been tried as well as tests with or without the spectrogram masked section (Fig. 6).

The evaluation and comparison of these models is based on five key metrics: Accuracy, balanced accuracy, precision, recall, and the F1 score. Each test also generates the associated Receiver Operating Characteristic (ROC) curve and Area under the ROC Curve (AUC) value for binary classifications.

1) PET results: Fig. 7 depicts the best PET classification model architecture. In addition to this architecture, the model uses the binary cross entropy loss function (BCELoss). An Adam optimizer with a 0.0001 learning rate and a OneCycle Learning Rate (OneCycleLR) scheduler, with a 0.001 maximum learning rate with a linear annealing strategy, are used.

This model presents good results on the testing set: an accuracy of 0.89, a balanced accuracy of 0.90, a precision of 0.84 and a recall of 0.95. The F1 score is 0.89.

As shown in Fig. 8, the classifier wrongly predicts PET too often, leading to errors when another type of waste is thrown away. But, the classifier predicts the Other class with only 5% errors. Since LusTra aims to educate young people, this is a good thing. It is not desirable to tell a child that a mistake was made when they put their PET bottle in the PET bin.

Fig. 9 shows the PET sound classification ROC. The model obtains an AUC of 0.96, meaning that it has a great capacity to distinguish between Other and PET.

2) Aluminium results: The aluminium classifier uses the same architecture as the PET one (Fig. 7): A better accuracy of 0.94, balanced accuracy of 0.91, a precision of 0.94, a worse recall of 0.85. The F1 score is the same as the PET one: 0.89.

Similar to PET, the AUC value is very good at 0.98 (Fig. 11), meaning that the system highly distinguishes the differences between Other and Alu.

Unlike PET, aluminium has a lower recall. Since the system is equipped with a metal detector that has better accuracy (around 99%), the classifier takes a secondary role. The

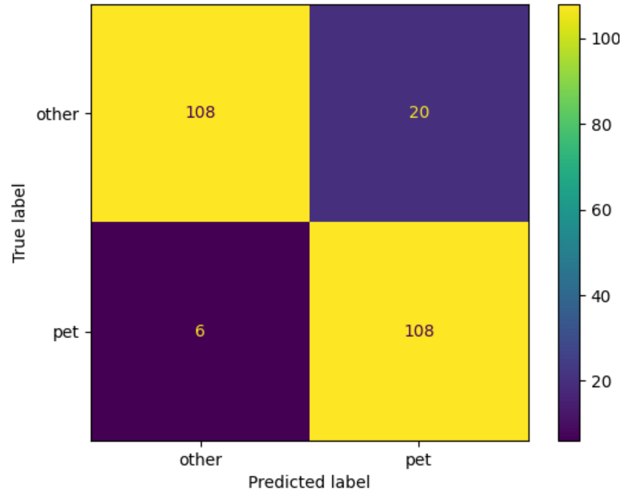


Fig. 8. PET sound classification confusion matrix

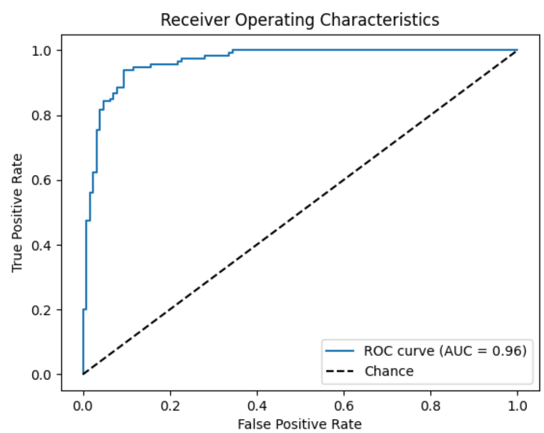


Fig. 9. PET sound classification ROC curve

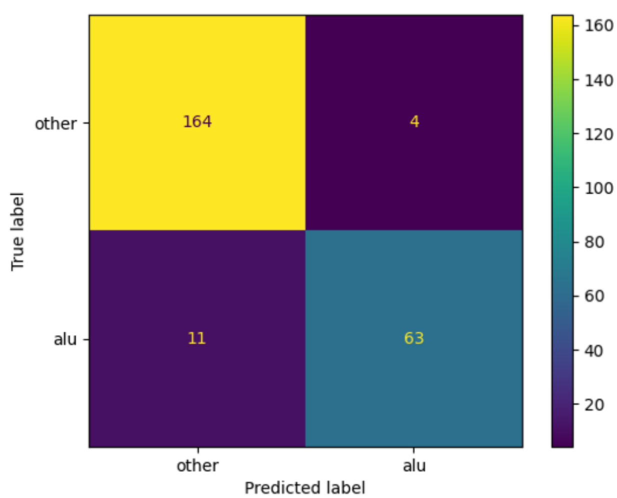


Fig. 10. Alu sound classification confusion matrix

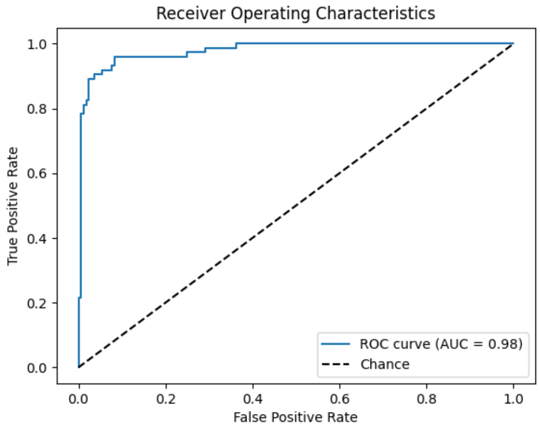


Fig. 11. Alu sound classification ROC curve

aluminium classifier is here as a fail-safe, in case the metal detector has an issue.

B. Accelerometer classification

Various approaches have been tried for the accelerometer:

1) Using a one dimensional convolution model;
2) Converting the data to a sound, following the same steps as the sound classification;
3) Extracting features from the signal and applying a random forest classification on them.

The first approach was quickly abandoned because the classifier rapidly overfitted during training. The same happened with the second one. The third approach provides the best results.

1) PET results: Predictions on the accelerometer classifier test dataset provide promising but not satisfactory results. With an accuracy of 0.76, this is a significant difference from the same prediction with sound (0.89). The other metrics follow the same trend, in parenthesis the difference with the sound:

- Balanced accuracy: 0.76 (-0.14);
- Precision: 0.81 (-0.03);
- Recall: 0.65 (-0.31);
- F1 score: 0.72 (-0.17);
- AUC: 0.87 (-0.06).

In Fig. 12 and 13, the PET confusion matrix and ROC curve are shown.

2) Aluminium results: Accelerometer results for aluminium prediction are better than the PET ones. They are satisfactory but still not as good as the sound one. The metrics are:

- Accuracy: 0.86 (-0.08)
- Balanced accuracy: 0.82 (-0.09)
- Precision: 0.76 (-0.18)
- Recall: 0.73 (-0.12)
- F1 score: 0.74 (-0.15)
- AUC: 0.88 (-0.10)

The confusion matrix (Fig. 14) confirms these results. In the ROC curve (Fig. 15), the accelerometer classifier is less discriminatory between the two classes than the sound equivalent.

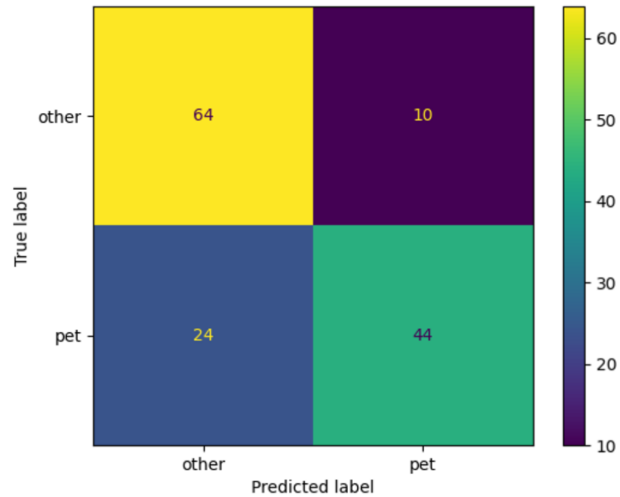


Fig. 12. PET accelerometer classification confusion matrix

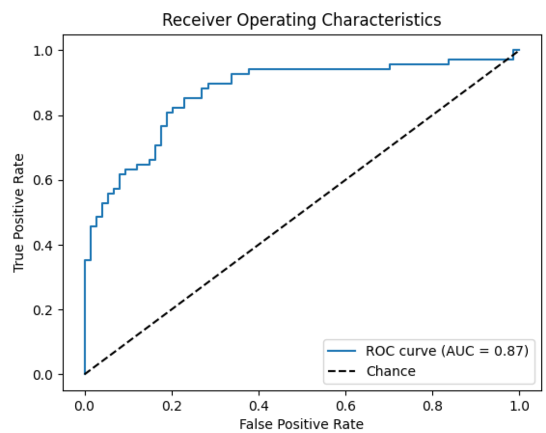


Fig. 13. PET accelerometer classification ROC curve

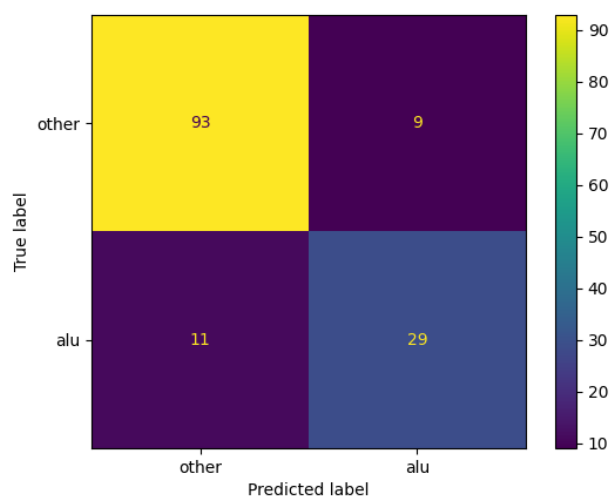


Fig. 14. Aluminium accelerometer classification confusion matrix

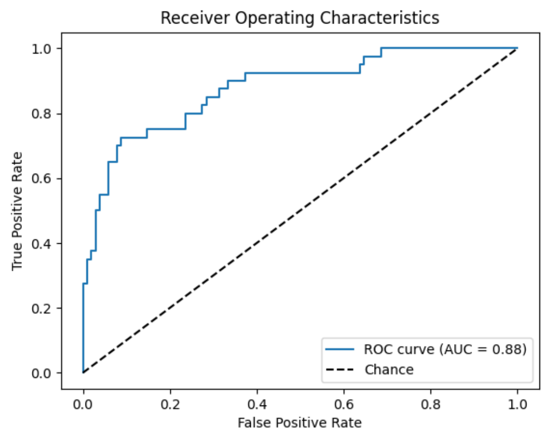


Fig. 15. Aluminium accelerometer classification ROC curve

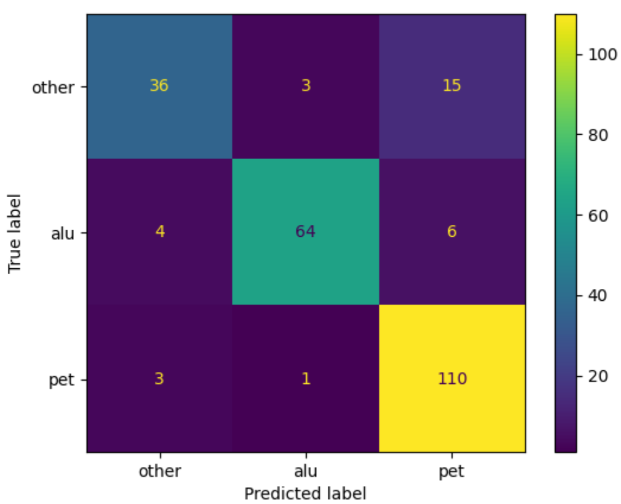


Fig. 16. Sound three classes classification confusion matrix

C. Three classes classification

Based on the results from the accelerometer classification, the decision was made to proceed using only sound data. The same architecture and parameters described in section V-A1 have been used with few alterations (to move from a binary classification to a multi-class classification): A cross-entropy loss function instead of a binary one; The sigmoid layer in the model (Fig. 7) has been removed.

The confusion matrix in Fig. 16 shows encouraging results for a multi-class problem. Like the binary PET classification, the PET predictions are more frequent when the expected value is Other. This can be explained by the distribution of each class in the collected data. With PET making up 46% of the collected data, the multi-class distribution is unbalanced. Especially with only 24% of "Other" values. The global accuracy of this classification is 0.87. It is less than the two binary classifications. Same for the balanced accuracy score at 0.83.

D. Data combination

Since accelerometer results are worse than sound results, an attempt has been made to combine the two types of signals (sound and accelerometer). Thus, the accelerometer signal is added as a new dimension to the sound. First, it is resampled (from 38.4kHz to 64kHz). Then, the new signal is concatenated to the sound resulting in a two-dimensional signal. This last signal is converted into a Mel spectrogram and the same operations as the binary sound classification (section V-A) are performed. The model architecture has been modified to optimise the results based on the new data. Unfortunately, with an accuracy of 0.75 for PET prediction, this method was rejected. This confirms our initial intuition that the accelerometer adversely affects data quality.

VI. Discussions

The developed system integrates three ways of classifying waste: The metal detector; The sound classifier; The accelerometer classifier.

The aluminium bin relies mainly on the metal detector which results in a 99% trustworthy accuracy.

For the PET bin, the initial approach is to check whether the metal detector detected any metal. If not, the system asks the two classifiers for their predictions. If the two predictions are identical, their value is returned. If the two classifiers disagree, the more confident classifier prediction value (closest to 0 or 1) is returned. Since the sound prediction quality is significantly better (0.89 accuracy) than the accelerometer (0.76 accuracy), only the sound predicted value is used. Keeping the accelerometer prediction enabled us to detect other side effects.

The whole system with multiple bins (PET and aluminium) has been tested in real-life settings. These tests reveal that the sound classifier works as expected, but with a lower accuracy than theoretically calculated. Environmental noise does not affect waste classifications, but the throwing gesture does. The data, collected by only two people, lack information on certain throwing types.

The system prediction accuracy in real-life situations has been calculated at 70% for sound, i.e a 19% difference compared to the theoretical one (see section V-A1).

The accelerometer classifier faces a major issue. It performs poorly on systems other than the one used to collect the training data. Despite having multiple bins, each with its own testing system, the accelerometer classifier performance is inconsistent due to data volatility, which skews the result.

VII. Conclusion

LusTra offers a new acquisition and classification method to detect good or bad recycling behaviour at the very beginning of the waste management chain. Using sound and accelerometer signals reduces costs, power consumption, and facilitates the installation of the system in a bin.

Our findings indicate that sound prediction holds significant promise as a method for tackling this challenge. However, the effectiveness of this approach could be further enhanced by incorporating more diverse data, specifically from individuals exhibiting varied throwing patterns. While accelerometer prediction shows encouraging potential, it currently falls short of achieving the same level of quality as sound prediction. We believe that the development and refinement of this method could greatly benefit from a larger and more diverse dataset. Future research should focus on these areas to drive further advancements in this field.

References

[1] Mallory Szczepanski. *What Technology Advancements Mean for the Future of Recycling*. www.proquest.com/docview/2334732891/abstract/2922F368340B4030PQ.. Accessed: 2024-06-03.

[2] Merve Erkinay Ozdemir et al. "Applying machine learning approach in recycling". en. In: *J. Mater. Cycles Waste Manag.* (Feb. 2021).

[3] Weisheng Lu and Junjie Chen. "Computer vision for solid waste sorting: A critical review of academic research". In: *Waste Management* 142 (2022), pp. 29–43. ISSN: 0956-053X. DOI: https://doi.org/10.1016/j.wasman.2022.02.009. URL: https://www.sciencedirect.com/science/article/pii/S0956053X22000678.

[4] Janusz Bobulski and Mariusz Kubanek. "Deep learning for plastic waste classification system". en. In: *Appl. Comput. Intell. Soft Comput.* 2021 (May 2021), pp. 1–7.

[5] Sishir Bhandari. "Automatic Waste Sorting In Industrial Environments Via Machine Learning Approaches". MA thesis. Tampere University, 2020. URL: https://trepo.tuni.fi/handle/10024/123574.

[6] George E. Sakr et al. "Comparing deep learning and support vector machines for autonomous waste sorting". In: *2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*. 2016, pp. 207–212. DOI: 10.1109/IMCET.2016.7777453.

[7] Shirazush Salekin Chowdhury et al. "The design and implementation of an autonomous waste sorting machine using machine learning technique". In: *AIUB Journal of Science and Engineering (AJSE)* 19.3 (Mar. 2021), pp. 134–142.

[8] João Sousa, Ana Rebelo, and Jaime S. Cardoso. "Automation of Waste Sorting with Deep Learning". In: *2019 XV Workshop de Visão Computacional (WVC)*. 2019, pp. 43–48. DOI: 10.1109/WVC.2019.8876924.

[9] Xiangru Chen. "Machine learning approach for a circular economy with waste recycling in smart cities". In: *Energy Reports* 8 (2022), pp. 3127–3140. ISSN: 2352-4847. DOI: https://doi.org/10.1016/j.egyr.2022.01.193. URL: https://www.sciencedirect.com/science/article/pii/S2352484722001937.

[10] Swann Puig and Noria Foukia. "CleverTrash: An ML-based IoT system for waste sorting with continuous learning cycle". In: *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. 2022, pp. 1–9. DOI: 10.1109/ICECET55527.2022.9872943.

[11] Gang Lu et al. "One-dimensional convolutional neural networks for acoustic waste sorting". In: *Journal of Cleaner Production* 271 (2020), p. 122393. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2020.122393. URL: https://www.sciencedirect.com/science/article/pii/S0959652620324409.