Scientific Reports Title to see here

Alice Author[1,*], Bob Author[2], Christine Author[1,2,+], and Derek Author[2,+]

[1]Affiliation, department, city, postcode, country
[2]Affiliation, department, city, postcode, country
[*]corresponding.author@email.example
[+]these authors contributed equally to this work

ABSTRACT

Example Abstract. Abstract must not include subheadings or citations. Example Abstract. Abstract must not include subheadings or citations. Example Abstract. Abstract must not include subheadings or citations. Example Abstract. Abstract must not include subheadings or citations. Example Abstract. Abstract must not include subheadings or citations. Example Abstract. Abstract must not include subheadings or citations. Example Abstract. Abstract must not include subheadings or citations.

Please note: Abbreviations should be introduced at the first mention in the main text – no abbreviations lists. Suggested structure of main text (not enforced) is provided below.

Introduction

Our planet is drowning in plastic litter that can sneakily enter our body over time. An estimation showed that in 2010 at least 4.8 million tons of plastic litter has already entered our ocean[1]; a volume that will increase without an effective waste management plan[1,2]. Once in the wild, plastic can persist for decades as most types are resistant to natural degradation processes[3]. Environmental influences, however, can cause it to disintegrate into micron-sized particles commonly known as microplastics[4–8]. Almost invisible to the eye, they have now been detected on nearly every corner of our planet[8–13], in animals[14–16] and even in our food[17,18]. Since 2021, we also have the first evidence that microplastics are present in humans, when Ragusa et al. detected microplastics in the human placenta[19].

Plastic litter is highly diverse due to environmental influences and the sheer endless possibilities to produce plastic with desired material properties. Therefore, to evaluate their detrimental effects on animals and humans, we need a deeper insight on the samples origins[20–22]. Here, tools to detect and classify plastic litter at different stages play an indispensable role. Studies on plastic pollution commonly use Raman and Fourier transform infrared (FTIR) spectroscopy solutions to analyse plastic samples[23–26]. Both techniques, however, come with physical limitations[24,27,28] and hence, we only cover a subset of plastic waste types out there.

Most recently, Ornik et al.[29] used photoluminescence (PL) spectroscopy for plastic identification. By comparing the intensity ratios in the PL spectrum of different samples, they sucessfully distinguished plastics from non-plastic samples in the riverine and marine environment. Such an identification method, however, may not be reproducible since a measurement can change with different experimental factors such as hardware alignments, sample sites or even scientific experiences. On the other hand, it is impractical to capture and quantify these influences since it is impossible to account for all possible factors. Nevertheless, it raises the question if there are subsets in the spectra that can be used for the identification while being robust against experimental variations. One possible solution is to capture a part of these variations and integrate them in a spectral library. Once implemented, algorithms and mathematical models help unraveling the origins of the plastic sample.

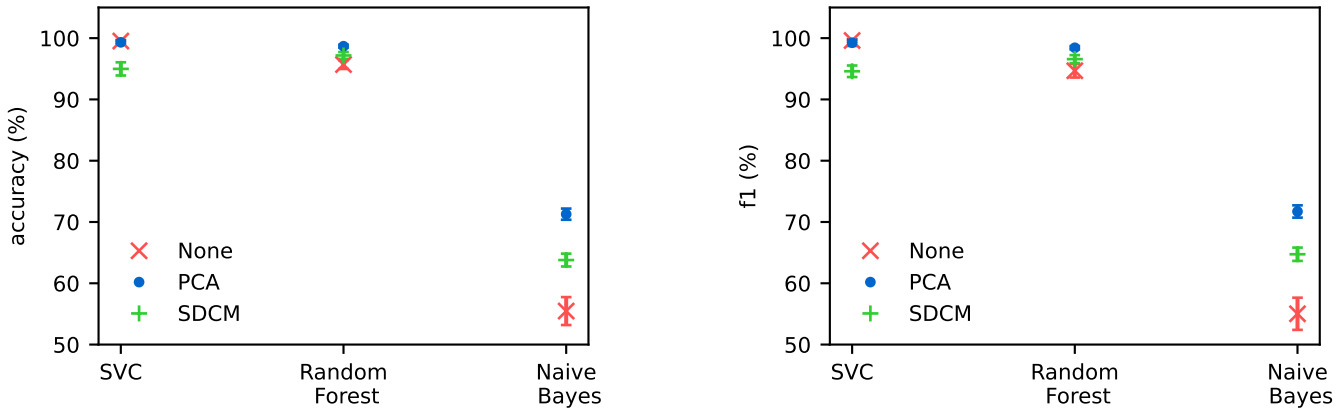
For predictions that account for data variations, machine learning (ML) models are suitable candidates. To generate a plastic waste prediction model, we apply a selected learning method on the labeled spectral dataset. The model's performance partially depends on the number of input variables in a dataset, which, in our case, is the measured intensity for each channel of our spectrometer. To improve the performance, it is common practice to reduce the number of variables while retaining the essential data information[30,31]. The efficiency to capture most of the information depends on the selected transformation technique. Here, recently published technique termed as signal dissection by correlation maximization (SDCM) stands out which successfully discovered new hidden signatures in a gene expression dataset[32]. This is particularly attractive for PL-based plastic litter detection as it can identify type specific subspectra that are robust against experimental variations.

In this report, we demonstrate that SDCM is suitable to generate robust PL-based plastic litter identification models. To demonstrate this, we look at two sets of ML models that are either based on a SDCM-transformed PL dataset or an untransformed

dataset. By comparing model's accuracies, i.e. the probability that a model based prediction is correct, we find consistently higher values for SDCM-based ML models than their counterparts. This underlines the robustness of SDCM-transformed PL datasets against experimental variations.

Results

Performance evaluation of ML models



(a) Accuracy

(b) f1

Figure 1. Plot of the performance metrics *accuracy* and *f1* for different plastic classification models.

We compare the performances between the classification models, by evaluating the performance metrics *accuracy* and *f1*. Figures 1a and 1b show plots of the values and standard deviations for the accuracy and f1, respectively. All values are also presented in the supplements in Tables 3a and 3b. The plot shows that the learning method has the biggest influence on the model's performance. Models generated with the SVC and the Random Forest algorithm achieve values over 90 %, while it drops by 20 % for models generated the Naive Bayes algorithm. Thus, for plastic classification the former two methods are likely to be more suitable in the future.

The dimension reduction technique, on the other hand, has little influence on the performance. SVC works best with the entire spectral data or when it is transformed with PCA as evidenced by values around 99 % for both the accuracy and f1. When applying SVC on SDCM-transformed data both metrics drop slightly by 5 %. For models generated with the Random Forest algorithm the values for accuracy and f1 are between 95 %–99 % and 94 %–99 %, respectively. Here, the algorithm works best with PCA-transformed data followed by SDCM-transformed data.

Classification performance of ML models

Next, we evaluate the performance of our models with respect to the different sample types in our dataset. Figure 2 presents a confusion matrix of all models in this study. It reveals that the performance for an individual class depends on the dimension reduction technique and learning method. For example, a model that uses random forest and SDCM-transformed data is better at identifying PA than a random forest model with PCA-transformed data. We also observe trends that are present in all models: first, PP gets mixed up as PE and second, PVC gets mixed up as a non-plastic material. These observations show, that more data is required so that the dimension reduction techniques can capture the spectral signatures to identify the classes.

Subsection

Example text under a subsection. Bulleted lists may be used where appropriate, e.g.

- First item
- Second item

Third-level section

Topical subheadings are allowed.

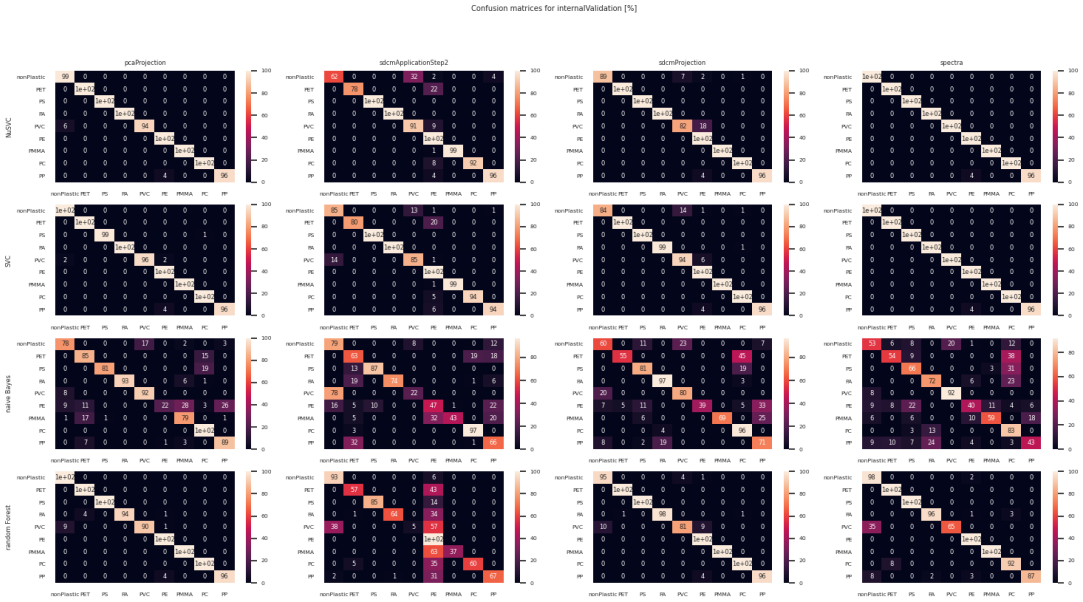


Figure 2. Confusion matrix for individual sample classes

Discussion

The Discussion should be succinct and must not contain subheadings.

Methods

Experimental setup

Figure 3 illustrates our experimental setup for PL spectroscopy measurements. The blue path highlights the incident beam that is used to excite the sample and induce photoluminescence. Our laser generates light with a central wavelength of 402 nm which passes through an excitation filter to select light with a wavelength of 405 nm. A dichroic mirror directs the light to lens 1 which focusses incident light on the sample's surface. The path that is taken by the emitted photoluminescence light is highlighted in red. Starting from the sample's surface, this light is collected and collimated by lens 1 and passes through the dichroic mirror. To ensure that the excitation light is completely removed from the emission path, we use a longpass filter with a cut-on wavelength of 420 nm. Finally, lens 2 focusses the light onto an optical fibre which directs the light to our spectrometer (LR2, Lasertack GmbH).

Both the laser and the spectrometer are controlled with a microcontroller which, in turn, is connected to a pc. This arrangement makes it possible to control the laser power, exposure time and the time between sample excitation and signal acquisition. The latter is set to 500 ms.

Samples and measurement parameters

Our PL spectra dataset is generated from 46 samples which consists of non-plastic materials from the riverine and marine environment and plastics from manufacturers and retail products. A summary of the dataset is presented in Table 2.

For each sample, we take two different measurements where we change the laser power P_{laser}, the exposure time t_{ex} and alignments in the setup. A list of these measurement parameters is presented in Table 1.

References

1. Jambeck, J. R. *et al.* Plastic waste inputs from land into the ocean. *Science* **347**, 768–771, DOI: 10.1126/science.1260352 (2015).

2. Geyer, R., Jambeck, J. R. & Law, K. L. Production, use, and fate of all plastics ever made. *Sci. Adv.* **3**, e1700782, DOI: 10.1126/sciadv.1700782 (2017).

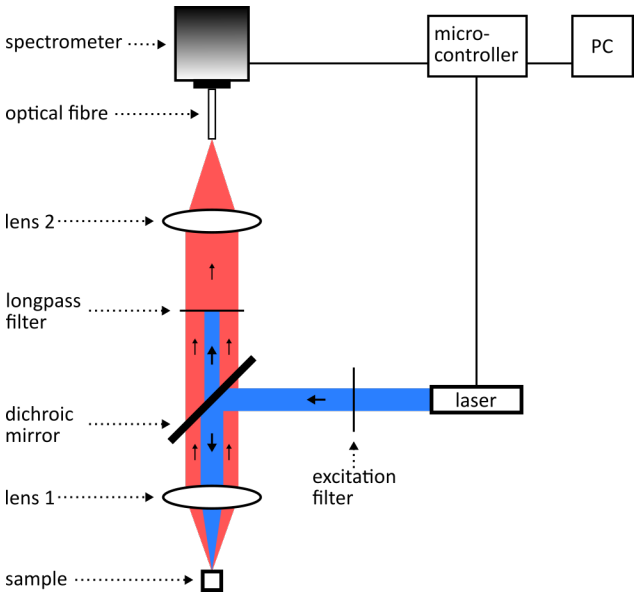


Figure 3. Illustration of the photoluminescence spectroscopy setup. The excitation light follows the path highlighted in blue to induce PL on the sample. The pathway of the PL signal is highlighted in red.

3. Chamas, A. *et al.* Degradation Rates of Plastics in the Environment. *ACS Sustain. Chem. & Eng.* **8**, 3494–3511, DOI: 10.1021/acssuschemeng.9b06635 (2020).
4. Thompson, R. C. Lost at Sea: Where Is All the Plastic? *Science* **304**, 838–838, DOI: 10.1126/science.1094559 (2004).
5. Julienne, F., Delorme, N. & Lagarde, F. From macroplastics to microplastics: Role of water in the fragmentation of polyethylene. *Chemosphere* **236**, 124409, DOI: 10.1016/j.chemosphere.2019.124409 (2019).
6. Zhang, K. *et al.* Understanding plastic degradation and microplastic formation in the environment: A review. *Environ. Pollut.* **274**, 116554, DOI: 10.1016/j.envpol.2021.116554 (2021).
7. Song, Y. K. *et al.* Combined Effects of UV Exposure Duration and Mechanical Abrasion on Microplastic Fragmentation by Polymer Type. *Environ. Sci. & Technol.* **51**, 4368–4376, DOI: 10.1021/acs.est.6b06155 (2017).
8. Duis, K. & Coors, A. Microplastics in the aquatic and terrestrial environment: sources (with a specific focus on personal care products), fate and effects. *Environ. Sci. Eur.* **28**, 2, DOI: 10.1186/s12302-015-0069-y (2016).
9. Chiba, S. *et al.* Human footprint in the abyss: 30 year records of deep-sea plastic debris. *Mar. Policy* **96**, 204–212, DOI: 10.1016/j.marpol.2018.03.022 (2018).
10. Napper, I. E. *et al.* Reaching New Heights in Plastic Pollution—Preliminary Findings of Microplastics on Mount Everest. *One Earth* **3**, 621–630, DOI: 10.1016/j.oneear.2020.10.020 (2020).
11. Eerkes-Medrano, D., Thompson, R. C. & Aldridge, D. C. Microplastics in freshwater systems: A review of the emerging threats, identification of knowledge gaps and prioritisation of research needs. *Water Res.* **75**, 63–82, DOI: 10.1016/j.watres.2015.02.012 (2015).
12. Andrady, A. L. Microplastics in the marine environment. *Mar. Pollut. Bull.* **62**, 1596–1605, DOI: 10.1016/j.marpolbul.2011.05.030 (2011).
13. Allen, S. *et al.* Atmospheric transport and deposition of microplastics in a remote mountain catchment. *Nat. Geosci.* **12**, 339–344, DOI: 10.1038/s41561-019-0335-5 (2019).
14. Barboza, L. G. A. *et al.* Microplastics in wild fish from North East Atlantic Ocean and its potential for causing neurotoxic effects, lipid oxidative damage, and human health risks associated with ingestion exposure. *Sci. The Total. Environ.* **717**, 134625, DOI: 10.1016/j.scitotenv.2019.134625 (2020).
15. Haave, M., Gomiero, A., Schönheit, J., Nilsen, H. & Olsen, A. B. Documentation of Microplastics in Tissues of Wild Coastal Animals. *Front. Environ. Sci.* **9**, DOI: 10.3389/fenvs.2021.575058 (2021).
16. Jamieson, A. J. *et al.* Microplastics and synthetic particles ingested by deep-sea amphipods in six of the deepest marine ecosystems on Earth. *Royal Soc. Open Sci.* **6**, 180667, DOI: 10.1098/rsos.180667 (2019).

Sample Category	No. of Samples	Sample Type
Non-plastic	12	• Sand • Wood • Posidonia Oceanica (Plant) • Sepia Officinalis (Bone) • Echinocardium Cordatum (Shell) • Hexaplex Eggs (Shell) • Monodonta Turbinata (Shell) • Neverita Josephina (Shell) • Lithophyllum Racemus
Plastic (manufacturer)	26	• Polyamide (PA) • Polycarbonate (PC) • Polyethylene (PE) • Low-density polyethylene (LDPE) • High-density polyethylene (HDPE) • Polyethylene terephthalate (PET) • Polymethylmethacrylate (PMMA) • Polypropylene (PP) • Polystyrene (PS) • Polyvinyl chloride (PVC)
Plastic (retail)	8	• LDPE • HDPE • PET • PP

Table 1. Overview of samples used for this study.

17. Koelmans, A. A. *et al.* Microplastics in freshwaters and drinking water: Critical review and assessment of data quality. *Water Res.* **155**, 410–422, DOI: 10.1016/j.watres.2019.02.054 (2019).

18. Zhang, Q. *et al.* A Review of Microplastics in Table Salt, Drinking Water, and Air: Direct Human Exposure. *Environ. Sci. & Technol.* **54**, 3740–3751, DOI: 10.1021/acs.est.9b04535 (2020).

19. Ragusa, A. *et al.* Plasticenta: First evidence of microplastics in human placenta. *Environ. Int.* **146**, 106274, DOI: 10.1016/j.envint.2020.106274 (2021).

20. Prata, J. C., da Costa, J. P., Lopes, I., Duarte, A. C. & Rocha-Santos, T. Environmental exposure to microplastics: An overview on possible human health effects. *Sci. The Total. Environ.* **702**, 134455, DOI: 10.1016/j.scitotenv.2019.134455 (2020).

21. Campanale, Massarelli, Savino, Locaputo & Uricchio. A Detailed Review Study on Potential Effects of Microplastics and Additives of Concern on Human Health. *Int. J. Environ. Res. Public Heal.* **17**, 1212, DOI: 10.3390/ijerph17041212 (2020).

22. Lim, X. Microplastics are everywhere — but are they harmful? *Nature* **593**, 22–25, DOI: 10.1038/d41586-021-01143-3 (2021).

23. Prata, J. C., da Costa, J. P., Duarte, A. C. & Rocha-Santos, T. Methods for sampling and detection of microplastics in water and sediment: A critical review. *TrAC Trends Anal. Chem.* **110**, 150–159, DOI: 10.1016/j.trac.2018.10.029 (2019).

24. Löder, M. G. J. & Gerdts, G. Methodology Used for the Detection and Identification of Microplastics—A Critical Appraisal. In *Marine Anthropogenic Litter*, 201–227, DOI: 10.1007/978-3-319-16510-3_8 (Springer International Publishing, Cham, 2015).

25. Sun, J., Dai, X., Wang, Q., van Loosdrecht, M. C. & Ni, B.-J. Microplastics in wastewater treatment plants: Detection, occurrence and removal. *Water Res.* **152**, 21–37, DOI: 10.1016/j.watres.2018.12.050 (2019).

Sample Category	Measurement 1		Measurement 2	
	P_{laser} [mW]	t_{ex} [ms]	P_{laser} [mW]	t_{ex} [ms]
Non-plastic	0.5–130	300	0.2–2.8	300
Plastic (manufacturer)	5–130	300	0.5–100	300
Plastic (retail)	0.5–130	300–1500	0.5–104	300

Table 2. Summary of samples and measurement parameters. For each sample, two measurements were taken where the laser power (P_{laser}), the exposure time (t_{ex}) and alignments in the setup were changed.

26. Zhang, Y. *et al.* Atmospheric microplastics: A review on current status and perspectives. *Earth-Science Rev.* **203**, 103118, DOI: 10.1016/j.earscirev.2020.103118 (2020).

27. Araujo, C. F., Nolasco, M. M., Ribeiro, A. M. & Ribeiro-Claro, P. J. Identification of microplastics using Raman spectroscopy: Latest developments and future prospects. *Water Res.* **142**, 426–440, DOI: 10.1016/j.watres.2018.05.060 (2018).

28. Xu, J.-L., Thomas, K. V., Luo, Z. & Gowen, A. A. FTIR and Raman imaging for microplastics analysis: State of the art, challenges and prospects. *TrAC Trends Anal. Chem.* **119**, 115629, DOI: 10.1016/j.trac.2019.115629 (2019).

29. Ornik, J. *et al.* Could photoluminescence spectroscopy be an alternative technique for the detection of microplastics? First experiments using a 405 nm laser for excitation. *Appl. Phys. B* **126**, 15, DOI: 10.1007/s00340-019-7360-3 (2020).

30. Aggarwal, C. C., Hinneburg, A. & Keim, D. A. On the Surprising Behavior of Distance Metrics in High Dimensional Space. 420–434, DOI: 10.1007/3-540-44503-X_27 (2001).

31. Fodor, I. K. A Survey of Dimension Reduction Techniques. Tech. Rep., Lawrence Livermore National Laboratory (LLNL), Livermore, CA (2002). DOI: 10.2172/15002155.

32. Grau, M., Lenz, G. & Lenz, P. Dissection of gene expression datasets into clinically relevant interaction signatures via high-dimensional correlation maximization. *Nat. Commun.* **10**, 5417, DOI: 10.1038/s41467-019-12713-5 (2019).

Acknowledgements (not compulsory)

Acknowledgements should be brief, and should not include thanks to anonymous referees and editors, or effusive comments. Grant or contribution numbers may be acknowledged.

Author contributions statement

Must include all authors, identified by initials, for example: A.A. conceived the experiment(s), A.A. and B.A. conducted the experiment(s), C.A. and D.A. analysed the results. All authors reviewed the manuscript.

Additional information

To include, in this order: **Accession codes** (where applicable); **Competing interests** (mandatory statement).

The corresponding author is responsible for submitting a competing interests statement on behalf of all authors of the paper. This statement must be included in the submitted article file.

Supplementary information

Calculated classification model metrics

Learning Method	None		PCA		SDCM	
	acc [%]	Δacc [%]	acc [%]	Δacc [%]	acc [%]	Δacc [%]
SVC	99.5	0.2	99.3	0.3	95.0	1.1
Random Forest	95.7	0.7	98.7	0.2	97.2	0.6
Naive Bayes	55.5	2.3	71.3	0.9	63.8	1

(a) Accuracy

Learning Method	None		PCA		SDCM	
	f1 [%]	Δf1 [%]	f1 [%]	Δf1 [%]	f1 [%]	Δf1 [%]
SVC	99.6	0.2	99.2	0.5	94.6	0.9
Random Forest	94.7	1.1	98.4	0.3	96.6	0.7
Naive Bayes	55.0	2.6	71.7	1	64.7	1.1

(b) f1

Table 3. Summary of samples and measurement parameters.