

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jes

JES
JOURNAL OF
ENVIRONMENTAL
SCIENCES
www.jesc.ac.cn

Rapid detection of colored and colorless macro- and micro-plastics in complex environment via near-infrared spectroscopy and machine learning

Hui-Huang Zou¹, Pin-Jing He^{1,2}, Wei Peng^{1,2}, Dong-Ying Lan¹,
Hao-Yang Xian¹, Fan Lü^{1,2}, Hua Zhang^{1,2,*}

¹Institute of Waste Treatment & Reclamation, College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China

²Shanghai Institute of Pollution Control and Ecological Security, Shanghai 200092, China

ARTICLE INFO

Article history:

Received 15 September 2023

Revised 25 November 2023

Accepted 3 December 2023

Available online 8 December 2023

Keywords:

Colorless microplastics

Near-infrared hyperspectral imaging

Plastic identification

Partial least squares discriminant analysis

Machine learning

ABSTRACT

To better understand the migration behavior of plastic fragments in the environment, development of rapid non-destructive methods for in-situ identification and characterization of plastic fragments is necessary. However, most of the studies had focused only on colored plastic fragments, ignoring colorless plastic fragments and the effects of different environmental media (backgrounds), thus underestimating their abundance. To address this issue, the present study used near-infrared spectroscopy to compare the identification of colored and colorless plastic fragments based on partial least squares-discriminant analysis (PLS-DA), extreme gradient boost, support vector machine and random forest classifier. The effects of polymer color, type, thickness, and background on the plastic fragments classification were evaluated. PLS-DA presented the best and most stable outcome, with higher robustness and lower misclassification rate. All models frequently misinterpreted colorless plastic fragments and its background when the fragment thickness was less than 0.1mm. A two-stage modeling method, which first distinguishes the plastic types and then identifies colorless plastic fragments that had been misclassified as background, was proposed. The method presented an accuracy higher than 99% in different backgrounds. In summary, this study developed a novel method for rapid and synchronous identification of colored and colorless plastic fragments under complex environmental backgrounds.

© 2024 The Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences. Published by Elsevier B.V.

Introduction

Plastic, including a wide range of organic polymers, is the most widely used material owing to its properties such as lightweight, high durability, and low cost (Adarsh et al., 2022b). However, plastics abuse and mismanagement have

caused incessant accumulation of waste plastics in the environment worldwide (Yakovenko et al., 2020). These plastic wastes can break down into smaller fragments (macroplastics and microplastics (MPs)) under the influence of different degradation factors (Yang et al., 2022b). When compared with large plastic fragments, MPs have size less than 5 mm, have greater migration capacity, and are a potential threat

* Corresponding author.

E-mail: zhanghua_tj@tongji.edu.cn (H. Zhang).

(Yu and Hu, 2022). Nevertheless, both macroplastics and MPs pose great challenges to different environments worldwide.

MPs have been confirmed to be an accessible vector to transport organic contaminants, heavy metals such as cadmium, polychlorinated biphenyls, and polycyclic aromatic hydrocarbons into living organisms through food chains, thus threatening the ecosystems (Gillibert et al., 2019). Hence, quantitative and qualitative detection of MPs are important to analyze their migration behavior and potential toxicity (Zhang et al., 2019). However, MPs identification is complex, involving pretreatment process and selection of suitable methods depending on their particle size (Yang et al., 2022a). In traditional MPs identification method, which is both time-consuming and labor-intensive, the samples are first observed under optical microscopy and then the particles are picked up for spectroscopic analysis. To overcome these limitations, novel MPs identification methods have been explored, such as pyrolysis gas chromatography/mass spectrometry (PY-GC/MS) (Xu et al., 2022), attenuated total reflectance–Fourier transform infrared spectroscopy (ATR-FTIR) (Morgado et al., 2021), laser-induced breakdown spectroscopy (LIBS) (Sommer et al., 2021), and Raman spectroscopy (Schymanski et al., 2018). However, PY-GC/MS can be destructive to the samples, LIBS shows potential safety concerns of creating a plasma on plastic, ATR-FTIR can obtain spectral information only in a small range of size (μm) (Michel et al., 2020), and Raman spectroscopy can be easily disturbed by the fluorescence background caused by inorganic substances, organic substances, and colored additives (Ye et al., 2022). Besides, complex sample preparation procedures, such as separation and size reduction, are still needed for these methods owing to small quantity of samples used for analysis. In addition, ATR-FTIR and Raman spectroscopy are suitable only for the detection of single type of MPs. In contrast, hyperspectral imaging (HSI) produces spectral and spatial signatures with abundant and interesting information, which can be extracted from the image pixels (Serranti et al., 2018). This method can be applied for the direct identification of MPs from various components based on their spectra, thus reducing unnecessary separation steps (Shan et al., 2018). Besides, HSI is an in-situ spectroscopy technique that does not damage the sample and can maintain the original shape for further analysis, with the identification procedure usually taking only a few minutes.

Spectroscopic techniques combined with machine learning model have been applied to the detection of MPs in different sites, such as intestinal tracts of fish (Zhang et al., 2019), soil (Li et al., 2021; Zhao et al., 2022), and marine environment (Huang et al., 2021; Vidal and Pasquini, 2021). However, the MPs need to be separated from the environmental samples before analysis, in order to avoid the interference of the complex environment media. Besides, in these previous studies, only limited plastic types or standard samples were used to develop the model, which decreased its generalization ability. The lack of spectral data of diverse plastic types and development of models without using actual plastic samples can inevitably lead to incorrect classification of some unknown samples into plastic types available in the database. It has been reported that significant misclassification occurred when using only standard polymers (Wu et al., 2020). Furthermore, while the influences of size, color, shape, and weathering factors on

the identification of MPs have been investigated (Huang et al., 2021; Phan et al., 2022; Vidal and Pasquini, 2021), the effect of different environments media (backgrounds) is usually ignored. Moreover, similar spectral information of environment media or substances may result in incorrect recognition and subsequent underestimation of colorless plastics (Wang et al., 2022a). Therefore, the influence of various environmental media should be taken into consideration during MPs identification, especially colorless MPs, in practical settings (Duan and Li, 2021; Sobhani et al., 2019), and the identification of test samples from different sources under different backgrounds based on comparison with the available data must be validated. However, only a few studies have paid attention to this important issue (Ye et al., 2022).

In the present study, a high-throughput detection method was developed by combining four machine learning models and HSI technique for the identification and classification of plastics in complex environment, including 15 types of macro-plastics, MPs, or rubbers, with organic and inorganic environmental backgrounds. The optimized model exhibits the capability to rapidly identify numerous types of colored plastics and rubbers across diverse environmental media with remarkable robustness, and without the need for sample preparation (extraction and purification). Subsequently, a two-stage identification method was applied for the identification of three commonly used colorless and thin plastic fragments that can be easily misidentified with the environment, which filled the gap in directly detecting colorless microplastic against varied environmental backgrounds. The developed method for synchronous and rapid detection of both colored and colorless plastic fragments in diverse environments provides a new insight for both qualitative and quantitative identification, as well as the management of plastic pollution.

1. Materials and methods

1.1. Sample collection and preparation

Compost and loess/quartz sand were used to represent the organic and inorganic environmental media, respectively. The compost samples (labeled as CP-1, CP-2, CP-3, CP-4, and CP-5) were collected from five food waste treatment plants (detailed information about the materials can be found in previous publication (Qu et al., 2022)). To eliminate potential interference of MPs and other impurities in the compost as much as possible, the plastics, metals, and glass were separated from the compost materials. The loess sample was collected from Lanzhou City, Gansu Province, China, and quartz sand sample was purchased from Sinopharm Chemical Reagent Co., Ltd.

The spectral data of CP-1, loess, quartz sand, and 51 types of virgin and actual samples comprising of 15 kinds of polymer materials placed on CP-1 were obtained for model development. The polymer materials include polyethylene (PE), polyethylene terephthalate (PET), polypropylene (PP), polystyrene (PS), polyvinyl chloride (PVC), polycarbonate (PC), polyamide (PA), polymethyl methacrylate (PMMA), thermoplastic polyurethane (TPU), natural rubber (NR), styrene butadiene rubber (SBR), cis-polybutadiene (PBD), nitrile butadiene

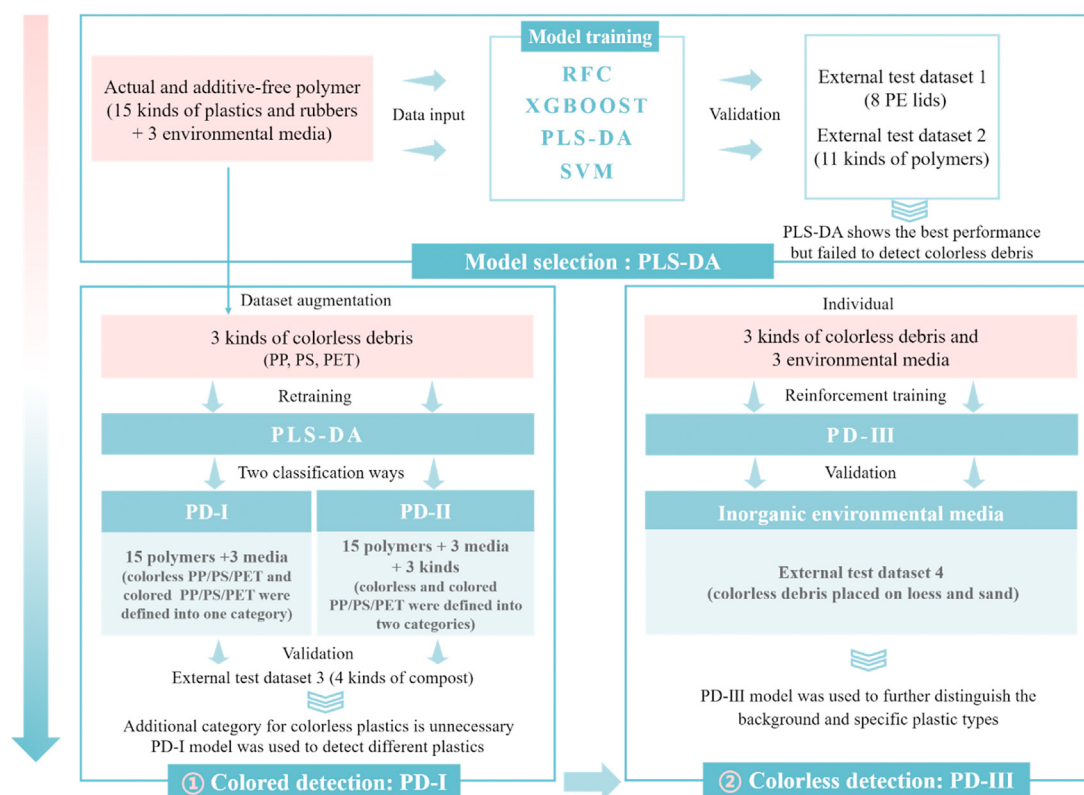


Fig. 1 – Architecture of identification for the four machine learning models.

rubber (NBR), ethylene propylene diene monomer (EPDM), and acrylonitrile-butadiene-styrene-copolymers (ABS). Additive-free samples of plastics and rubber (virgin samples, 2–4 mm wide) were purchased from Usolf Chemical Technology Company (Shandong, China).

Four external test datasets were prepared, in which the samples and data were not included in model training and test datasets. Eight PE lids from PET bottles with different colors (white, yellow, orange, red, pink, purple, green, and blue) were used as external test dataset 1 to examine the influence of color on MPs detection by the four development models. 23 actual waste samples of 11 kinds of polymer materials were placed on CP-1 to obtain their spectral data (Appendix A Fig. S1), including PP clip, PP caps of centrifugal tube, PS food package, PS soap box, PE cap of drinking bottle, PET drinking bottle, PET health product bottle, PC compact disc, PA nylon rope, PA gasket, PMMA plastic disk, ABS lego block, nitrile rubber gloves, PVC gloves, and NR block. They serve as external test dataset 2 to validate the efficacy of the four developed models and to determine the optimal model for subsequent advancements. In addition, the actual colorless PP, PS, and PET samples were cut into square fragments with a width of 4–10 mm, and then placed on the organic (CP-2, CP-3, CP-4, CP-5) and inorganic environmental media (loess and quartz sand) to obtain external test dataset 3 and 4, respectively. The thickness of PP, PS, and PET fragments were 0.03–0.76, 0.10–0.36, and 0.30–0.46 mm, respectively. Each type of colorless samples comprised 10 pieces for external test dataset 3 and 4, which were used to explore the impact of environmental media to validate the efficacy of PD-I and PD-III, respectively. The archi-

tecture of identification for the four machine learning models can be seen in Fig. 1.

1.2. Hyperspectral image acquisition

For hyperspectral image acquisition, all the samples were placed on the composts or loess and quartz sand. The spectrum of three environmental media were directly obtained through HSI system. The HSI tests were performed using a near-infrared (NIR) system (FX17, Specim, Finland) with wavelength ranging from 935 to 1720 nm. The HSI system was equipped with a halogen tungsten lamp to serve as the light source, and the analytical process was conducted in a black box to avoid ambient light interference. A field of view of 38° was adopted. The distance between the sample and the lens is 260 mm. Based on this setup, the theoretical minimum recognition size of the instrument is about 280 μm . The exposure time and frame rate were set at 3.8 msec and 50 Hz, respectively, which were optimized prior to the experiment. To guarantee the stability of image acquisition, each experiment was conducted in triplicate.

1.3. Data processing

The spectral information of each sample was checked with ENVI software to ensure high-quality data acquisition, which is crucial for model development. Then, regions of interest (ROI) were selected for each hyperspectral image to remove any irrelevant pixels and retain the most representative data for polymer particles. The ROI containing three-

dimensional data were unfolded in the spatial direction to two-dimensional matrices. For each kind, 1000 data (1000 pixels) were randomly chosen from the entire ROI of different samples. A database comprising 15 kinds of polymers, a black polymer category (black polymers were classified as a separate category), and three environmental media (CP-1, loess, and quartz sand) was established and serve as training and testing datasets (containing 19,000 data) for the development of four models. For each modeling, the whole dataset was randomly sorted three times and then split into training (70%) and test (30%) datasets. Five-fold cross-validation was used to divide the training dataset into five folds, training subset into four folds, and validation subset into one fold. The average value was used to present the final result. To ensure correct identification, suitable libraries are necessary and must not only include spectra of virgin samples, but also of actual plastics as well as non-plastic materials in the environment (Ivleva, 2021).

Principal component analysis (PCA) was applied using SPSS 21 to reduce the dimensionality of the additive-free polymers and analyze the similarity in the spectra. Based on Varimax and Kaiser normalization method, two principal components were extracted.

1.4. Machine learning methods

Supervised classification methods, including random forest classifier (RFC), extreme gradient boost (XGBOOST), partial least squares-discriminant analysis (PLS-DA), and support vector machine (SVM) were chosen to identify polymers among all environmental surroundings. XGBOOST is a boosting model, which ensembles decision trees to minimize the error with a gradient descent optimization algorithm. PLS-DA is a multivariate classification method based on the PLS regression model (da Silva et al., 2020). SVM is a widely used supervised learning algorithm for classification tasks, which converts the data into a high-dimensional space with the kernel function. The classification results were assessed based on accuracy, recall, precision, and F1_score (Lan et al., 2023a).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1_score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

where TP, TN, FP, and FN are true positive numbers, true negative numbers, false positive numbers, and false negative numbers, respectively. Accuracy indicates the proportion of correctly predicted samples to the total samples; the higher the accuracy value is, the better is the classification result. Precision is the ratio of correctly predicted positive samples among all the samples classified as positive. Recall denotes the correctly identified positive samples among all true positive samples. F1_score is an indicator that comprehensively considers precision and recall; the higher the F1_score is, the better is the classifier.

2. Results and discussion

2.1. Spectral data analysis of virgin and actual samples

Appendix A Fig. S2 shows the spectra of 15 kinds of standard polymers, 3 kinds of actual plastics debris and 5 kinds of compost materials from ROI with the minimum, mean, and maximum values. PCA was applied to visualize the internal relationship of various polymers as well as verify the rationality of classification. As indicated in Fig. 2a, the factor scores for the two principal components suggested that most of the samples could be effectively classified. Although the plots exhibited “polymer-type clustered” structure, inner differences could still be observed owing to surface irregularities, different penetration depth of the electromagnetic radiation, and resolution (Amigo et al., 2015). The clusters of LDPE and HDPE samples were close and overlapped (Fig. 2b), and were both identified as PE owing to their similar composition. With respect to colorless and thin plastics on CP-1, it was difficult to distinguish between colorless PP debris and compost (CP-1) as environmental media, whereas colorless PET debris showed a

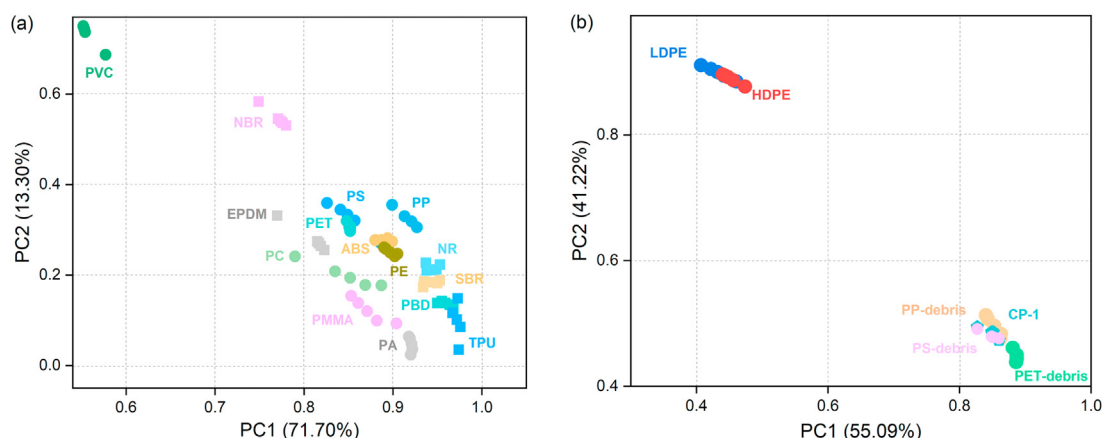


Fig. 2 – PCA score plots of the spectral data for (a) virgin samples with fifteen kinds of materials and (b) both virgin and actual samples with six types of materials.

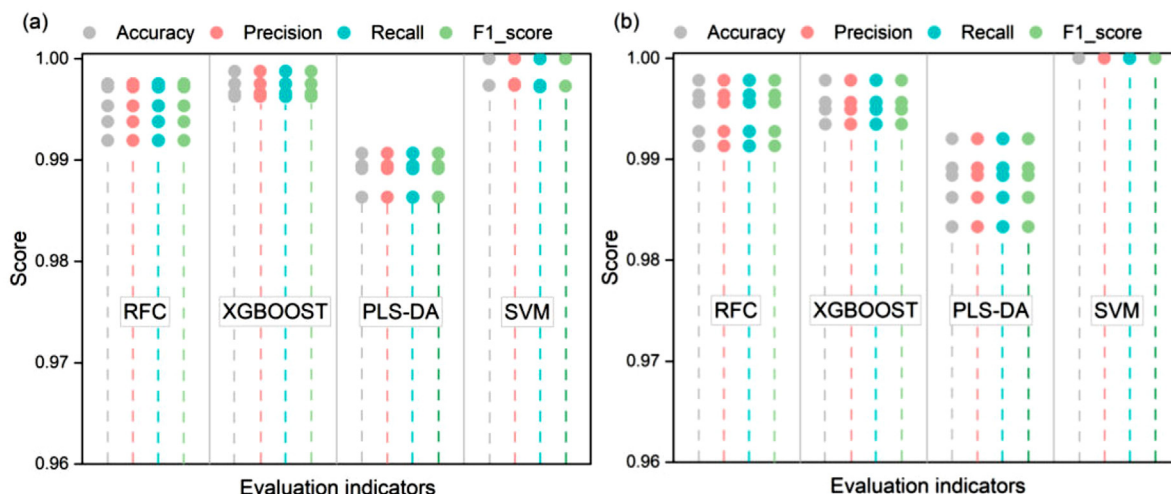


Fig. 3 – Five-fold cross-validation results of the four evaluation indicators for the four models with (a) training dataset and (b) test dataset.

spectral peak at around 1650 nm, which was different from the background and thus could be distinguished.

2.2. Model performance evaluation

Four model evaluation indicators, including accuracy, precision, recall, and F1_score, were calculated for assessment of the classification results (Fig. 3). The dots of the same color are the results from the five-fold cross-validation. Irrespective of training or test dataset, the RFC, XGBOOST, PLS-DA, and SVM models showed insignificant differences in the scores and consistently excellent performance with indicator values >98%. These results confirmed that the four models could accurately extract the polymer features of different components from the data, thus achieving excellent prediction results for classification, which could be attributed to significant simi-

larity in the distribution between training and test datasets. Similar findings have also been reported in a previous study on electrical and electronic equipment waste classification (Wu et al., 2020). Thus, external test dataset with independent samples that had not been used for model development is necessary for the validation of trained models.

2.3. Validation of model performance on external test dataset

2.3.1. Cross-transitivity verification of the model in terms of colors

As shown in Appendix A Fig. S3 and Fig. 4a, the pixels of all samples with different colors were almost correctly classified as PE, except for one cap with black area in the middle, which was categorized as “black polymer” instead of PE.

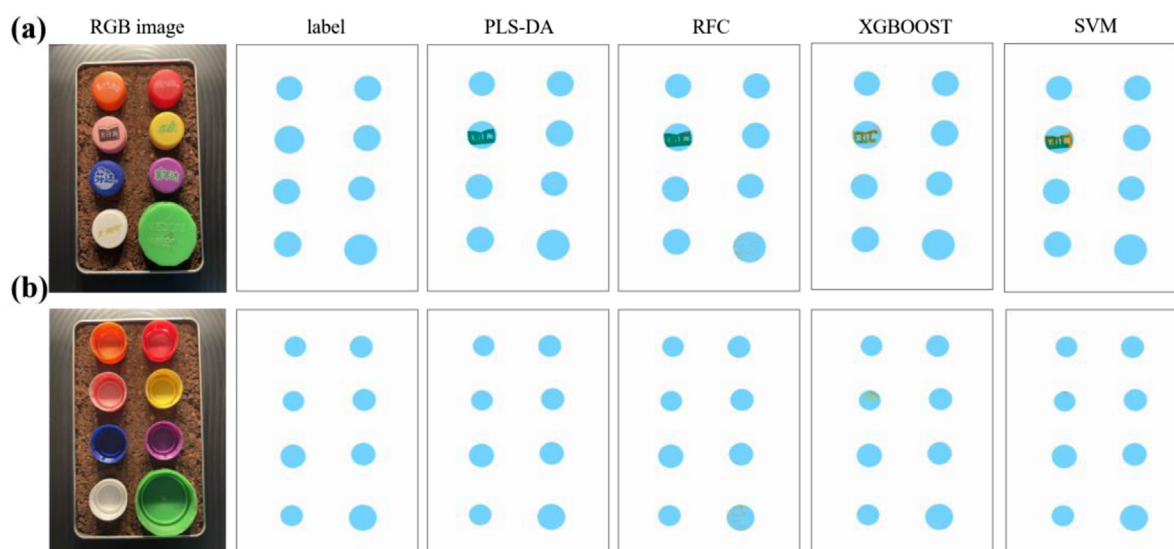


Fig. 4 – Samples, labels, and their prediction results obtained using the four models. (a) Outer and (b) inner parts of the PE bottle cap.

This result indicated that the classification was based on polymer type instead of color or shape, implying that small quantity of PE additives such as colorants and fillers did not significantly alter the spectral characteristics of the plastics. However, the specific material of black polymer could not be identified owing to strong absorption by carbon black within current wavelength range used in this study. Hence, further research is needed to address the problems associated with identification of black samples. Moreover, RFC, XGBOOST, and SVM failed to address the “edge effect”, which affected the identification of pixels between black and pink areas in the samples; in contrast, PLS-DA exhibited more even results.

Some points/pixels, unevenly distributed within the cap, were erroneously classified by RFC and XGBOOST (Fig. 4b). For instance, about 270 and 500 pixels inside the reusable caps were misclassified as EPDM by RFC and XGBOOST, respectively, which can be attributed to surface wear and signal noise resulting from varied roughness. In other words, alteration in the light intensity received by one pixel changed the light intensity reflected to the detector (Amigo et al., 2015). Besides, the spectral similarity between EPDM and HDPE also contributed to misclassification; however, the effect was lim-

ited owing to the good performance achieved by the training and test datasets. In contrast, PLS-DA could overcome spectral data heterogeneity between training and external test datasets, presenting the best performance in solving surface wear of polymers. In addition, PLS-DA was less sensitive to edge effects among uneven spectra and poorly focused regions of particles, similar to that reported by da Silva et al. (2020), and the few misclassifications could be well resolved by voting method (Jacquin et al., 2021).

2.3.2. Verification of model on actual samples

The spectra of actual samples that were not included in the training model were used as the external test dataset. The household plastic waste samples were classified into 11 kinds of polymers (Appendix A Fig. S1), as shown in Fig. 5a, b and Appendix A Fig. S4. The PLS-DA model presented the best classification results in terms of total accuracy and uniformity, which could be mainly owing to the fact that spectra data was compressed first and the variables with smaller dimensions were subsequently projected into a new space for classification. The training dataset used for modeling and test dataset used for evaluation followed similar feature distribution, which allowed promising model predictions. Misclassification of pixels

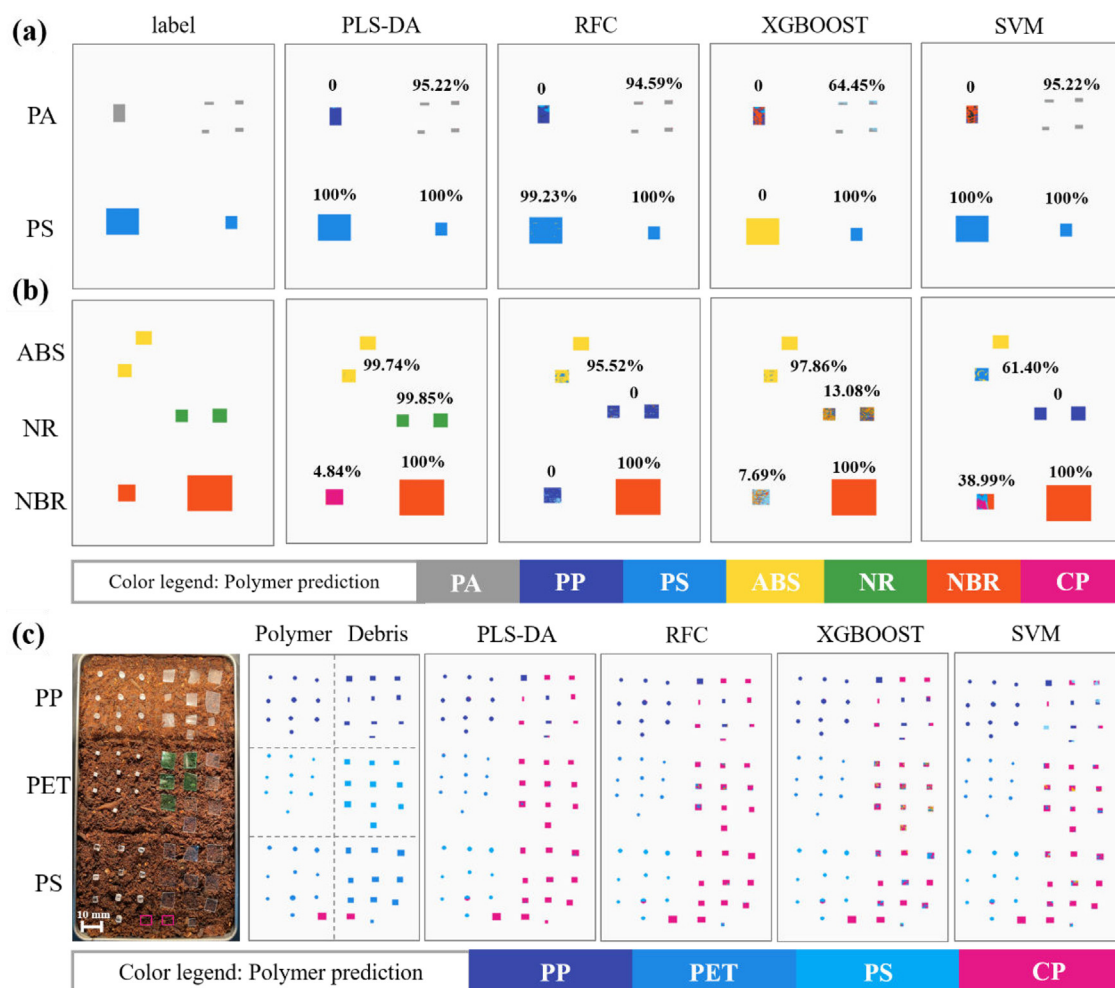


Fig. 5 – Labels of varied household waste samples and their prediction results using the four models. (a) PA and PS; (b) ABS, NR, and NBR; (c) Fragments of three additive-free polymers (left) and actual debris (right).

in the same sample was evident in the prediction results generated by RFC and XGBOOST models. Although HSI has the advantage of obtaining the spectral data of each pixel, it also has the disadvantage of presenting uneven prediction. The polymer surface and spectra obtained by hyperspectral cameras are seldom totally flat, which can promote unevenness. To solve this problem, voting method can be adopted to improve the classification results (Lan et al., 2023b). In the voting method, if the accuracy of individual sample (one ROI) is above 0.5, then the pixels are classified based on the majority results. In other words, the majority pixel results determine the polymer type, while the minority pixel results that are usually misclassified are discarded, which prevents the interference of uneven surface of the sample.

In addition to the spectral diversity of the same material between datasets, other factors may also affect object recognition. For instance, the shadow caused by a certain depth of the building block structure made of ABS (Fig. 5b) and the slight reflection produced by surface bending enhanced the spectral difference between the actual sample and reference material, which increased the difficulty of model identification. While RFC and SVM models were easily affected by these factors, PLS-DA model could overcome these problems and demonstrated excellent prediction ability with consistent color in one ROI. In particular, the PLS-DA model presented 100% and more than 99.7% accuracy rate for PS and ABS, respectively, which are higher than those reported in a previous study by Wu et al. (2020) (6.27% and 6.6% misclassification of PS and ABS, respectively) in which the classification model was developed based on NIR spectra of actual plastics adopting four spectral data preprocessing methods. Furthermore, the PLS-DA model achieved 96.68% and 100% accuracy rate for PET and PC in the present study, which are higher than those reported in the study by Duan and Li (2021) (12.9% and 71.9% for household PET and PC plastics, respectively).

Although PLS-DA model presented the best prediction ability, it also revealed abnormal colors in one label for PA (Fig. 5a) and NR (Fig. 5b). Moreover, PLS-DA model could correctly identify white nitrile gloves and PA plastic, but failed to detect blue nitrile gloves and PA rope. These results indicated the presence of a gap in the spectra between the actual samples and standard samples, leading to incorrect classification. This spectral difference could have been caused by the presence of a variety of additives in the actual samples, different surface conditions, and processing technology employed to determine the texture and durability functions. Furthermore, the collected wastes, despite mainly consisting of the same polymer, varied in their composition and surface properties; consequently, their spectral characteristics differed between training and test datasets. As the plastic additives introduced undesired signals, the actual plastics presented irregular adsorption spectra, when compared with the standard polymers (Zhao et al., 2022). This issue cannot be resolved by only adjusting and optimizing the model parameters, and can be rectified by adding the spectral variables of the actual samples to extend the model database as much as possible for model development, so that the model can learn the spectral characteristics of polymers with different additives and surface properties in diverse backgrounds to optimize its generalization ability.

As shown in Fig. 5c, PLS-DA model exhibited good prediction ability for both additive-free polymers and some actual samples, but showed significant classification error for colorless debris. This finding indicated that common model training methods are not suitable for the identification of colorless samples, because the spectra of colorless samples and background usually overlap and these models do not consider the interactive effect of mixed spectra. As electromagnetic radiation has certain penetrability, a pixel spectrum may not only contain information about the surface layer of the colorless sample, but may also include data about the nearest inner layers. Thus, the obtained spectra of the colorless samples may be a superposition of the sample itself and the background. Accordingly, to identify colorless samples in the environment, two strategies were proposed. In both these methods, the compound spectra of colorless PP/PS/PET samples and environmental background (CP-1, loess, and quartz sand) were respectively used for model development. However, in the first strategy, the spectral information of colorless PP/PS/PET was supplemented in the PP/PS/PET category and trained as the same column with colored ones. The colorless PP and colored PP were defined as one category for modeling. In contrast, in the second strategy, the colorless and colored samples were used as two categories for model development, and then combined to present the results for PP/PS/PET. The colorless PP and colored PP were defined as two categories for modeling. Subsequently, PLS-DA was chosen for comparison, and the new PLS-DA models trained using these methods were denoted as PD-I and PD-II, respectively. As shown in Fig. 6, both the models could identify almost all additive-free polymers as well as most of the colorless and actual plastic samples. While all the PET samples could be ascertained, some PP and PS samples were incorrectly identified, which attributed to variations in thickness and transmittance among the actual samples. Overall, the classification of colorless microplastics into a new category for training was unnecessary because it did not improve the accuracy of the model.

2.4. Colorless MPs identification in different environmental media

2.4.1. Universal applicability of the model in organic environmental media

To explore the generality of the developed model and its potential application to practical environments, varied organic backgrounds were included in the present study for plastic fragments identification. Owing to the optimal performance of the PLS-DA model, PD-I method was chosen to conduct the prediction for external test dataset. Fig. 7 shows the prediction results under four different compost backgrounds (CP-2, CP-3, CP-4, and CP-5), with the concrete values summarized in Appendix A Table S1. PLS-DA presented good prediction results with accuracies of 0.9239, 0.8731, 0.8607, and 0.9428, respectively, which further proved the effectiveness of the developed model. The high accuracies indicated that PLS-DA could cover spectral variety within the class (da Silva et al., 2020), which can be explained by the nature of the model. PLS-DA model uses the difference among classes to determine the regression coefficient, thus defining the class feature as a multivariate space. PLS-DA regression coefficient comprehensively

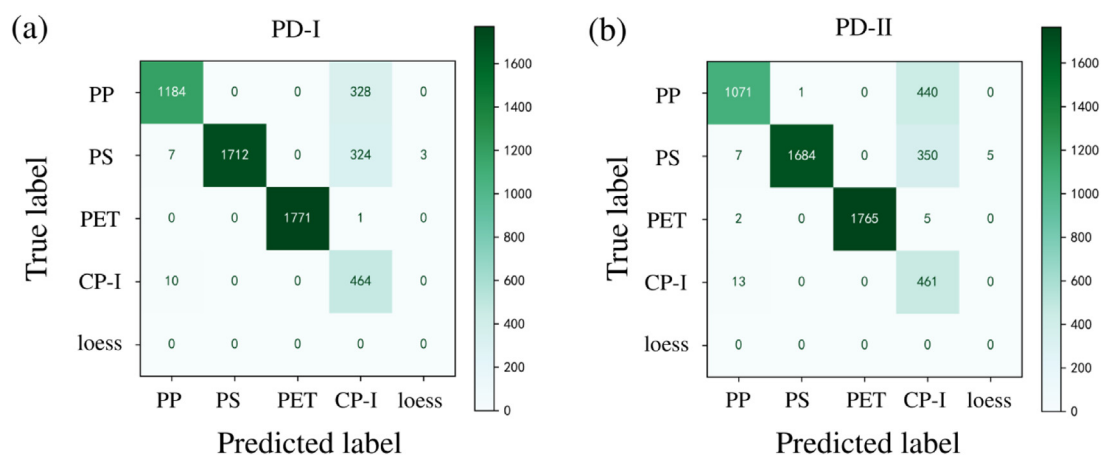


Fig. 6 – Classification of additive-free polymers and actual plastic debris using the two label classification methods: (a) PD-I (merged label of colored and colorless samples) and (b) PD-II (separate labels of colored and colorless samples).

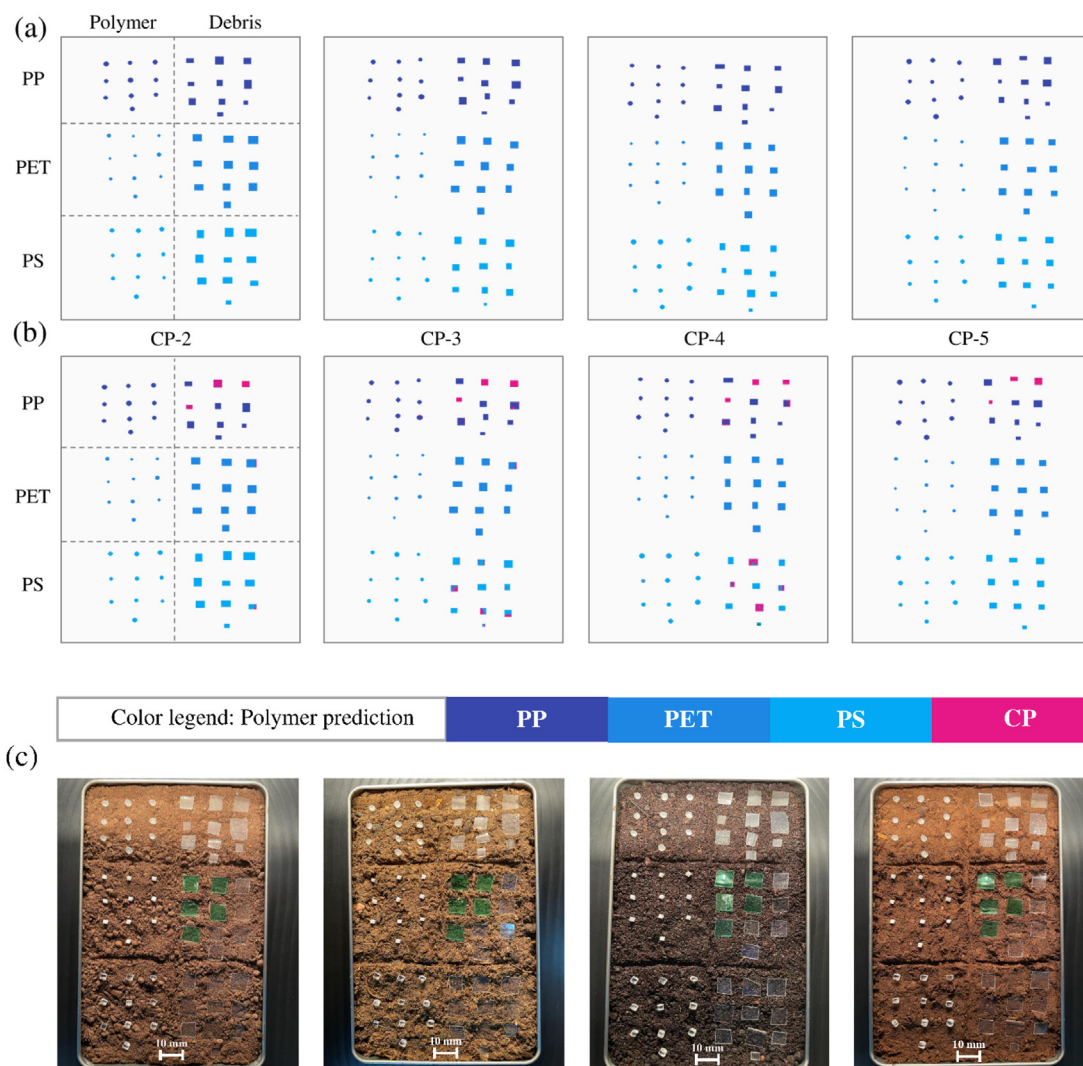


Fig. 7 – Classification of additive-free polymers and actual plastic debris in the four different compost materials by PLS-DA model: (a) true label and (b) predicted label.

considers spectral information, which can describe category variability and maximize its differences with other categories. The regression coefficients are orthogonal to each other and describe the unique information about the class. Therefore, PLS-DA model can better identify the minor spectral changes caused by particle characteristics (roughness) and edge effects, as indicated earlier (see Section 2.3). Moreover, it must be noted that the accurate identification of pixels is closely related to the similarity of backgrounds covered in model development to practical application scenarios.

The detection of virgin samples was almost independent of the background, and the varied accuracies were mainly owing to the differences in the identification of actual samples. With regard to actual samples placed on CP-2, the background appeared to exert only a slight impact on the identification of PET and PS, and only three PP fragments were wrongly identified as background. This error may be owing to varied thickness of the plastics; the thickness of incorrectly identified PP fragments was about 30–40 μm , while that of correctly identified PP fragments was about 260–760 μm . As the PET samples were cut from beverage bottles, they tended to be thicker than the other samples in the range of 300–460 μm . Interestingly, although the thickness of PS fragments (110–120 μm) was similar to that of PP fragments, the PS fragments were well distinguished from the background. It can be observed that the spectral bands of the PS fragments within 1650 and 1720 nm were different from those of the background, which may contain important information for classification, and that the model could augment these differences by compressing the characteristic spectra to distinguish them from the background (Amigo et al., 2015). Compost materials are rich in organic matter, which produces spectral peaks similar to those of PP fragments. Consequently, the spectral peaks of very thin PP fragments overlapped with those of organic matter, resulting in misidentification between the sample and background

by the model. PLS-DA model showed the worst recognition accuracy for PS in CP-4 compost owing to the higher difference between CP-4 and CP-1 used for model development, when compared with the other compost materials (Appendix A Fig. S2). As a result, PS fragments in CP-4 were identified as new samples, different from those learned by model development, causing significant errors. In summary, the thick plastic samples retained more intrinsic spectral information, whereas the thin samples showed more spectral information of the background. Consequently, the model identified colorless thin samples as the compost background.

2.4.2. Universal applicability of the developed model in inorganic environmental media

Wang et al. (2022b) found that the model developed using the data of organic soil for predicting soil organic carbon content exhibited huge prediction deviation when applied for the prediction of mineral soil. Therefore, in the present study, quartz sand and loess were used to determine the impact of inorganic backgrounds on the developed model.

Concrete prediction results based on pixels under two inorganic environmental media can be seen in Appendix A Table S2. As shown in Fig. 8, the classification of additive-free polymers was less affected by the inorganic background, whereas different kinds of actual samples present varied classification results. The PS fragments with thickness around 100 μm were more susceptible to the influence of inorganic background, resulting in edge recognition errors. Furthermore, the PP fragments with thickness around 30 μm were completely misidentified, whereas those with thickness of more than 700 μm were correctly detected. This is because the spectral information of very thin colorless samples overlapped with that of the background to produce a new spectrum, which was unknown to the trained model. Consequently, the model was unable to classify these samples accurately according to the

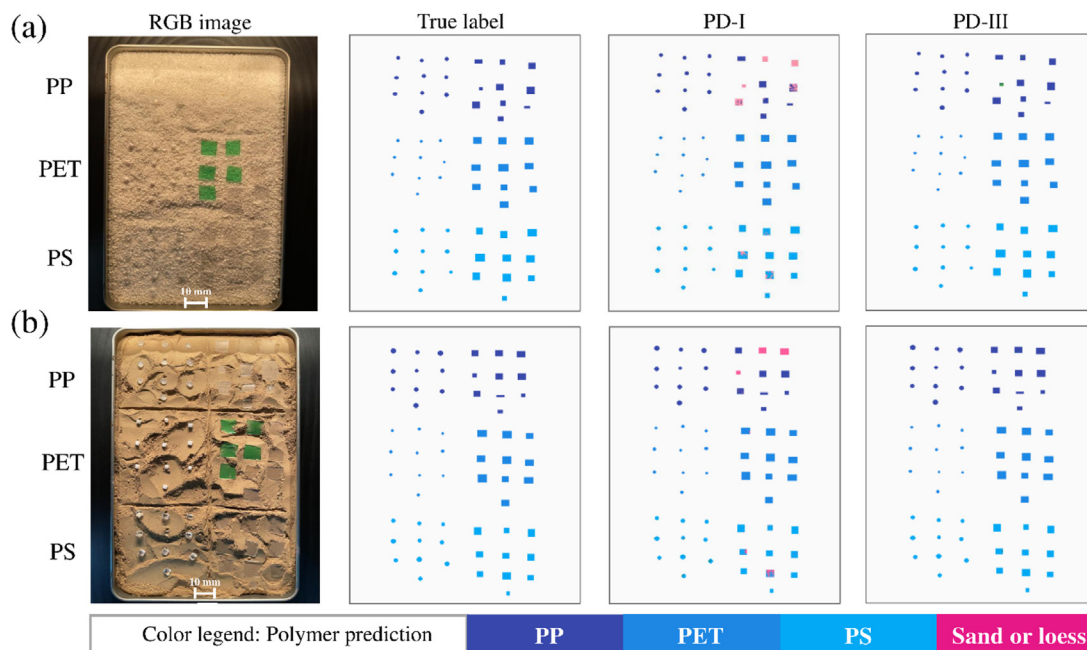


Fig. 8 – Classification of additive-free polymers and actual plastic debris in (a) sand and (b) loess backgrounds by the two PLS-DA models.

polymer characteristics in the training dataset, resulting in significantly random classification. In contrast, irrespective of the loess or sand background, almost all PET fragments were accurately identified. Thus, thickness may be one of the important factors affecting model identification, and the appropriate sample thickness for effective detection of samples may vary among different plastic types owing to the presence of diverse additive contents (Adarsh et al., 2022a; Signoret et al., 2020).

To identify the PP fragments, a two-stage identification model was proposed: PD-I model was used to detect different plastics and PD-III model was applied to samples identified as the background to further distinguish the background and specific plastic types. Only the spectra of three actual plastic debris and additive-free polymers in different backgrounds (CP-1, loess, or sand) were used for developing the PD-III model (Fig. 8), and the samples used for training and the plastic fragments to be predicted were not the same. The PP fragments with thickness around 30µm were accurately identified by the PD-III model, processing developed based on the features of those backgrounds and might not be well associated with the learned polymer spectral signatures available in the database, and may only identify the samples according to the nearest known features. Thus, further studies are needed to develop models covering a wider range of backgrounds.

2.5. Future prospect

In this study, the impact of different spectral pretreatment methods on the model performance was not discussed. To improve the model performance in practical applications, more samples from different sources and regions should be included in the training dataset and the model must be enabled to completely learn the features of different polymers. Besides, the influence of surface coating, surface contamination, additives, and pretreatment on plastic identification should be further investigated. Hyperspectral unmixing could be an effective way to solve these problems. A two-stage model classification method is necessary to overcome the decrease in the accuracy of the identification models when the difference between the actual sample and standard sample is significant, because the spectra of contaminants or additives might superimpose those of the samples to generate new spectra.

3. Conclusion

In this study, a novel method combining NIR-based HSI technology and machine learning classification models (RFC, XG-BOOST, PLS-DA, and SVM) was proposed, which could achieve rapid and accurate identification of plastics in different practical backgrounds. The PLS-DA model showed the best and most stable classification performance. All the four models presented good prediction ability for different actual samples, when the background influence was not taken into consideration. The color of polymer surface, except black color, had limited effect on model classification. However, the classification of colorless and thin plastics was significantly affected by the background. To avoid background interference, a two-stage identification method was developed to identify color-

less plastics in different backgrounds. In summary, this study provides an effective strategy for high-throughput detection and characterization of colored and colorless plastics, which could prevent underestimation of plastics pollution owing to misclassification as background and could be crucial for plastic pollution control.

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

This work was supported by the National Natural Science Foundation of China (No. 22276139), and the Shanghai's Municipal State-owned Assets Supervision and Administration Commission (No. 2022028).

Appendix A Supplementary data

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jes.2023.12.004.

REFERENCES

- Adarsh, U.K., Bhoje Gowd, E., Bankapur, A., Kartha, V.B., Chidangil, S., Unnikrishnan, V.K., 2022a. Development of an inter-confirmatory plastic characterization system using spectroscopic techniques for waste management. *Waste Manag.* 150, 339–351.
- Adarsh, U.K., Kartha, V.B., Santhosh, C., Unnikrishnan, V.K., 2022b. Spectroscopy: a promising tool for plastic waste management. *Trends Anal. Chem.* 149, 116534.
- Amigo, J.M., Babamoradi, H., Elcoroaristizabal, S., 2015. Hyperspectral image analysis. A tutorial. *Anal. Chim. Acta* 896, 34–51.
- da Silva, V.H., Murphy, F., Amigo, J.M., Stedmon, C., Strand, J., 2020. Classification and quantification of microplastics (<100 µm) using a focal plane array-fourier transform infrared imaging system and machine learning. *Anal. Chem.* 92 (20), 13724–13733.
- Duan, Q., Li, J., 2021. Classification of common household plastic wastes combining multiple methods based on near-infrared spectroscopy. *ACS ES&T Eng.* 1 (7), 1065–1073.
- Gillibert, R., Balakrishnan, G., Deshoules, Q., Tardivel, M., Magazzù, A., Donato, M.G., et al., 2019. Raman tweezers for small microplastics and nanoplastics identification in seawater. *Environ. Sci. Technol.* 53 (15), 9003–9013.
- Huang, H., Sun, Z., Liu, S., Di, Y., Xu, J., Liu, C., et al., 2021. Underwater hyperspectral imaging for in situ underwater microplastic detection. *Sci. Total Environ.* 776, 145960.
- Ivleva, N.P., 2021. Chemical analysis of microplastics and nanoplastics: challenges, advanced methods, and perspectives. *Chem. Rev.* 121 (19), 11886–11936.
- Jacquín, L., Imoussaten, A., Troussat, F., Perrin, D., Montmain, J., 2021. Control of waste fragment sorting process based on MIR imaging coupled with cautious classification. *Resour. Conserv. Recy.* 168, 105258.

- Lan, D., He, P., Qi, Y., Wu, T., Xian, H., Wang, R., et al., 2023a. Optimizing the quality of machine learning for identifying the share of biogenic and fossil carbon in solid waste. *Anal. Chem.* 95 (9), 4412–4420.
- Lan, D., He, P., Qi, Y., Wu, T., Xian, H., Wang, R., et al., 2023b. Machine learning and hyperspectral imaging-aided forecast for the share of biogenic and fossil carbon in solid waste. *ACS Sustain. Chem. Eng.* 11 (10), 4020–4029.
- Li, Y., Yao, J., Nie, P., Feng, X., Liu, J., 2021. An effective method for the rapid detection of microplastics in soil. *Chemosphere* 276, 128696.
- Michel, A.P.M., Morrison, A.E., Preston, V.L., Marx, C.T., Colson, B.C., White, H.K., 2020. Rapid identification of marine plastic debris via spectroscopic techniques and machine learning classifiers. *Environ. Sci. Technol.* 54 (17), 10630–10637.
- Morgado, V., Gomes, L., Bettencourt da Silva, R.J.N., Palma, C., 2021. Validated spreadsheet for the identification of PE, PET, PP and PS microplastics by micro-ATR-FTIR spectra with known uncertainty. *Talanta* 234, 122624.
- Phan, S., Padilla Gamiño, J.L., Luscombe, C.K., 2022. The effect of weathering environments on microplastic chemical identification with Raman and IR spectroscopy: part I. polyethylene and polypropylene. *Polym. Test.* 116, 107752.
- Qu, Q., Zhang, N., Hu, T., Zhao, T., Cui, G., 2022. Agricultural and forestry application potential of food waste fertilize produced by machine and the evaluation of secondary pollutants. *Environ. Sanit. Eng.* 02 (30), 31–36.
- Schymanski, D., Goldbeck, C., Humpf, H.U., Fürst, P., 2018. Analysis of microplastics in water by micro-Raman spectroscopy: release of plastic particles from different packaging into mineral water. *Water Res.* 129, 154–162.
- Serranti, S., Palmieri, R., Bonifazi, G., Cózar, A., 2018. Characterization of microplastic litter from oceans by an innovative approach based on hyperspectral imaging. *Waste Manag.* 76, 117–125.
- Shan, J., Zhao, J., Liu, L., Zhang, Y., Wang, X., Wu, F., 2018. A novel way to rapidly monitor microplastics in soil by hyperspectral imaging technology and chemometrics. *Environ. Pollut.* 238, 121–129.
- Signoret, C., Caro Bretelle, A.S., Lopez Cuesta, J.M., Ienny, P., Perrin, D., 2020. Alterations of plastics spectra in MIR and the potential impacts on identification towards recycling. *Resour. Conserv. Recycl.* 161, 104980.
- Sobhani, Z., Al Amin, M., Naidu, R., Megharaj, M., Fang, C., 2019. Identification and visualisation of microplastics by Raman mapping. *Anal. Chim. Acta* 1077, 191–199.
- Sommer, C., Schneider, L.M., Nguyen, J., Prume, J.A., Lautze, K., Koch, M., 2021. Identifying microplastic litter with laser induced breakdown spectroscopy: a first approach. *Mar. Pollut. Bull.* 171, 112789.
- Vidal, C., Pasquini, C., 2021. A comprehensive and fast microplastics identification based on near-infrared hyperspectral imaging (HSI-NIR) and chemometrics. *Environ. Pollut.* 285, 117251.
- Wang, Q., Guan, C., Han, J., Chai, M., Li, R., 2022a. Microplastics in China Sea: analysis, status, source, and fate. *Sci. Total Environ.* 803, 149887.
- Wang, S., Guan, K., Zhang, C., Lee, D., Margenot, A.J., Ge, Y., et al., 2022b. Using soil library hyperspectral reflectance and machine learning to predict soil organic carbon: assessing potential of airborne and spaceborne optical soil sensing. *Remote Sens. Environ.* 271, 112914.
- Wu, X., Li, J., Yao, L., Xu, Z., 2020. Auto-sorting commonly recovered plastics from waste household appliances and electronics using near-infrared spectroscopy. *J. Clean. Prod.* 246, 118732.
- Xu, Y., Ou, Q., Jiao, M., Liu, G., van der Hoek, J.P., 2022. Identification and quantification of nanoplastics in surface water and groundwater by pyrolysis gas chromatography–mass spectrometry. *Environ. Sci. Technol.* 56 (8), 4988–4997.
- Yakovenko, N., Carvalho, A., ter Halle, A., 2020. Emerging use thermo-analytical method coupled with mass spectrometry for the quantification of micro(nano)plastics in environmental samples. *Trends Anal. Chem.* 131, 115979.
- Yang, Z., Lü, F., Hu, T., Xu, X., Zhang, H., Shao, L., et al., 2022a. Occurrence of macroplastics and microplastics in biogenic waste digestate: effects of depackaging at source and dewatering process. *Waste Manag.* 154, 252–259.
- Yang, Z., Lü, F., Zhang, H., Wang, W., Xu, X., Shao, L., et al., 2022b. A neglected transport of plastic debris to cities from farmland in remote arid regions. *Sci. Total Environ.* 807, 150982.
- Ye, Y., Yu, K., Zhao, Y., 2022. The development and application of advanced analytical methods in microplastics contamination detection: a critical review. *Sci. Total Environ.* 818, 151851.
- Yu, F., Hu, X., 2022. Machine learning may accelerate the recognition and control of microplastic pollution: future prospects. *J. Hazard. Mater.* 432, 128730.
- Zhang, Y., Wang, X., Shan, J., Zhao, J., Zhang, W., Liu, L., et al., 2019. Hyperspectral imaging based method for rapid detection of microplastics in the intestinal tracts of fish. *Environ. Sci. Technol.* 53 (9), 5151–5158.
- Zhao, S., Zhang, Y., Qiu, Z., He, Y., Zhang, Y., 2022. Towards a fast and generalized microplastic quantification method in soil using terahertz spectroscopy. *Sci. Total Environ.* 841, 156624.