Trends in Analytical Chemistry 166 (2023) 117221

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

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

Issues with the detection and classification of microplastics in marine sediments with chemical imaging and machine learning

Reaha Goyetche [a], Leire Kortazar [a,b], José Manuel Amigo [a,c,*]

[a] Department of Analytical Chemistry, University of the Basque Country UPV/EHU, P.O. Box 644, 48080, Bilbao, Basque Country, Spain
[b] Plentzia Marine Station, University of the Basque Country (PiE-UPV/EHU), Plentzia-Bizkaia, Basque Country, Spain
[c] IKERBASQUE, Basque Foundation for Science, 48009, Bilbao, Spain

ARTICLE INFO

Keywords:
Microplastics
Sand
Hyperspectral imaging
NIR
Classification
Machine learning

ABSTRACT

Numerous studies have attempted to detect microplastic litter directly in environmental sediments via spectral imaging and powerful classification algorithms. Spectral imaging is attractive largely due to the benefits of adding a spatial element to spectral data, the relative measuring speed, and minimal sample processing. Despite this promise, important concerns related to the spatial and spectral selectivity must be considered along with the appropriateness of classification algorithms. Here we evaluate the performance of near infrared hyperspectral imaging (NIR-HSI) and four commonly used classification algorithms on a simple test case in which images of individual microplastics of known size on top of sand were collected. The results highlight major weak points of NIR-HSI and machine learning as applied to the detection of the microplastics, with a large proportion of false positives and negatives in most of the situations studied, and alerts the reader to important concerns about the use of this methodology.

1. Introduction

Microplastics (MPs) were defined as all plastic particles smaller than 5 mm in size, based on their potential for ingestion by biota during the International Research Workshop on the Occurrence, Effects and Fate of Microplastic Marine Debris hosted by NOAA in 2008 [1]. The diameter, surface area, and shape of MPs influence their potential for transport and bioavailability. This, in turn, influences the fate and transport of persistent, bioaccumulative, and toxic substances (PBTs) that may adsorb to plastics [2], such as persistent organic pollutants (POPs), polycyclic aromatic hydrocarbons (PAHs), or metals [3–6]. The lack of standardized analytical and quantification techniques in MP research makes comparison of the findings difficult [7]. Analysis to detect MPs in sand or sediment tends to follow these steps: 1) extraction from the matrix, such as sieving or density separation; 2) isolation and quantification; and 3) characterization and identification [8]. However, this methodology is time-consuming and can lead to material loss, especially for polymers denser than common separation solutions [9].

Some analytical techniques allow for quantification of MPs without characterization of polymer type. Visual analysis may be employed to locate plastic particles as small as 0.25 mm–2 mm, but identification varies greatly between observers and is highly subject to error [10]. Several destructive techniques may also be used to characterize polymers, including thermogravimetric analysis [11] and pyrolysis-gas-chromatography/mass spectrometry [12]. Stereomicroscopy can analyze plastic particles in the micron order of magnitude and provide physical characteristics and abundance [13]. The most common techniques for characterization of the polymer type are FT-IR (Fourier-transform infrared spectroscopy) in attenuated total reflectance (ATR) mode [9,14] and Raman spectroscopy [9,15], which are non-destructive techniques.

Due to the numerous analytical techniques available, near-infrared hyperspectral imaging (NIR-HSI) has been somewhat overlooked as a means of characterizing and quantifying MPs until recently, despite its application for plastic sorting in recycling facilities. Near-infrared (NIR) spectroscopy is a vibrational spectroscopic technique based on overtones and combination bands of molecular vibrations in the electromagnetic region between 780 and 2500 nm [16]. This leads to greater band overlap between compounds, which starkly contrasts with the "spectral fingerprint" and chemical specificity of Raman and FT-IR spectroscopy [17]. With NIR-HSI, meanwhile, a NIR spectrum in a full wavelength range is collected for each pixel in an image, adding a valuable spatial dimension to the spectral data [18]. In contrast with FT-IR and Raman spectroscopy, NIR-HSI is relatively faster and does not

* Corresponding author. Department of Analytical Chemistry, University of the Basque Country UPV/EHU, P.O. Box 644, 48080, Bilbao, Basque Country, Spain.
E-mail address: josemanuel.amigo@ehu.eus (J.M. Amigo).

https://doi.org/10.1016/j.trac.2023.117221
Received 2 June 2023; Received in revised form 14 July 2023; Accepted 2 August 2023
Available online 3 August 2023

require extensive sample preparation.

Given these benefits, NIR-HSI has emerged as a candidate for a fast, reliable, non-destructive technique to characterize MP waste, thereby expanding the possibilities for pollution monitoring [19–21]. The lack of certainty surrounding the detection limit for MP detection with NIR-HSI inhibits its widespread implementation. In one study, researchers collected MPs with surface-trawling plankton nets (200 μm mesh) and located fishing lines, but did not attempt to identify MP fragments smaller than 0.5-1 mm [20]. Meanwhile, Zhu et al. (2020) [22] reported a limit of detection of 100 μm using a modified NIR-HSI instrument to detect 11 polymer types, though they noted the limit of detection using the factory settings was 250 μm. In 2022, Piarulli et al. [19] reported a limit of detection, the calculation of which was based on pixel size, of 50 μm for detecting MPs in mussel digestate and filtered seawater. Another study developed an automatic detection and classification technique for MPs ranging from 150 μm to 5 mm in size directly in sand [21].

Because of differences in instrumentation, a single detection limit cannot be defined for all NIR-HSI cameras, but it is generally assumed to be at least the size of a pixel [21,23]. The difficulty in defining the edge of a MP contributes to the uncertainty surrounding the limit of detection for NIR-HSI [21]. The same problem was observed with micro-Fourier-transform infrared with a focal plane array, another hyperspectral imaging technique [24]. Indeed, recent monitoring efforts leverage the advantages of NIR-HSI with remote sensing for beached and floating marine litter [25,26]. At present, these remote sensing techniques are only valid for plastic litter greater than 25 mm in size, due to limitations that arise from the signal-to-noise ratio and concentration of plastics [27].

Advanced machine learning methodologies must be applied to glean meaningful results from the enormous quantities of data produced by NIR-HSI devices and allow the user to detect, quantify, and classify the elements in the scanned surface [28]. Classification is the most common chemometric strategy used for automatic detection and characterization of MPs [24]. Classification refers to various supervised multivariate analytical methods designed to sort samples into classes based on a series of measurements or theoretical values [29,30]. Validation evaluates the reliability of a model using training and validation sets comprised of objects for which the class is known and measurements or theoretical values have been defined.

Classification techniques fall into one of two primary categories: pure classification (discriminant) and class-modelling [29]. Pure classification techniques focus on differences between samples of different classes. All classes need to be defined with a corresponding set of standard spectra assigned unequivocally to a specific class. Three of the most commonly used pure classification methods are Partial Least Squares-Discriminant Analysis (PLS-DA), support vector machines (SVM), and Artificial Neural Networks (ANN). However, many polymer types may be present in a real sample, and each polymer class requires a corresponding spectral definition. As such, large databases are normally used to create a multiclass model to address this issue. Recent NIR spectroscopy and NIR-HSI studies have characterized MPs using PLS-DA [20,24,31], SVM [32–34], and convolutional neural networks [35]. Two recent studies claimed that convolutional neural networks were able to determine the presence of PP, polyethylene, and polyvinyl chloride in farmland soil using visible-NIR HSI [36,37], although there was no attempt to locate individual MPs within the soil. Class-modelling techniques partly overcome the aforementioned issues, since they are based on similarities among samples of the same class [30]. This allows the construction of models for individual classes [38]. Among these methods, Soft Independent Modelling by Class Analogy (SIMCA) stands out for its simplicity [39]. As a result, SIMCA is one of the most popular class-modelling techniques and was recently implemented to characterize MPs by polymer type [21].

Classification methods are pixel-based predictive models. That is, each pixel is predicted as belonging to a specific class without considering the adjacent pixels. That means each pixel is classified based only upon its spectrum and similarity to the spectra used to build the classification models. This is an important issue, as we will discuss further, since measuring pixels surrounded by environmental media may drastically affect the quality of the collected spectra due to light scattering and other morphological effects. Important analytical concerns must be investigated to understand the implications of using a non-selective spectroscopic technique—i.e. NIR-HSI—to find trace elements—i.e. MPs—in a complex environmental matrix. Addressing these concerns requires a deep understanding of the interactions between spectral radiation, MPs, environmental matrices, and the machine learning classification techniques.

This manuscript addresses the aforementioned concerns and sets a basis for understanding the extent to which NIR-HSI and machine learning can be used to detect MPs in complex matrices. The interest is centered on the growing trend of using NIR-HSI to detect and classify MPs directly in marine sediments like sand. However, most of the findings can be extrapolated to any detection protocol involving spectroscopy and machine learning. This manuscript hypothesizes that NIR-HSI, together with the proper classification method, is a viable technique to identify MPs and measure their size directly in sand, thereby avoiding time-consuming separation techniques. We present a simple case study to highlight some of the concerns regarding NIR-HSI as a technique for the detection and characterization of MPs. These concerns include the uncertain limit of detection, the difficulty in determining the edges of particles, and the inability to detect black particles. To this end, a simple case was tested. Standardized MPs of known polymeric identity were placed in standardized sand and NIR-HSI images were collected. This was done to assess the reliability of classification chemometric techniques of different complexity. The four classification methods evaluated were Partial Least Squares-Discriminant Analysis (PLS-DA), Soft Independent Modelling by Class Analogy (SIMCA), feed-forward artificial neural networks (ANN), and support vector machines (SVM).

2. Materials and methods

2.1. Polymers

Four different polymers were used: white acrylonitrile butadiene styrene (ABS), transparent polystyrene (PS), transparent polypropylene (PP), and black polybutylene terephthalate (PBT). These plastics were chosen from the list of polymers produced in the greatest quantities in Europe [40]. The plastics used in this study were kindly provided by the INNOSORT consortium (http://innosort.teknologisk.dk/) (further information in the following reference [41]). These polymers were in the form of pellets approximately 5 mm in diameter. The pellets contained no additives. MPs ranging from 200 μm to 1 mm in major axis length were manually cut from the pellets using a scalpel.

The size and Raman spectra of the MPs were measured with a DMLM Leica microscope (Renishaw, Gloucestershire, UK) coupled to an InVia confocal Raman spectrometer. The Raman spectra of the MPs (Fig. 1) were measured using a 785 nm excitation laser 10 accumulations of 5 s with 15 mW laser power to confirm polymer identity. Still images of each MP were collected with 20x magnification (Fig. 1). The area and the major axis length of the MPs were calculated by selecting three different thresholds and calculating the mean and the standard deviation (Table 1).

2.2. Sample preparation

The sand used in this study was collected from the surface layer of sand (2 cm. approx.) at a local beach in Plentzia, Spain (43°24′40″N 2°56′46″W). The sand was rinsed with tap water and floating debris was removed. The sand was then dried overnight at 100 °C. To remove any remaining traces of organic material or plausible MPs present, the sand was heat treated in an oven at 450 °C for 1 h. The sand was stored in a glass container. Petri dishes were filled with sand to a depth of 0.5 cm.

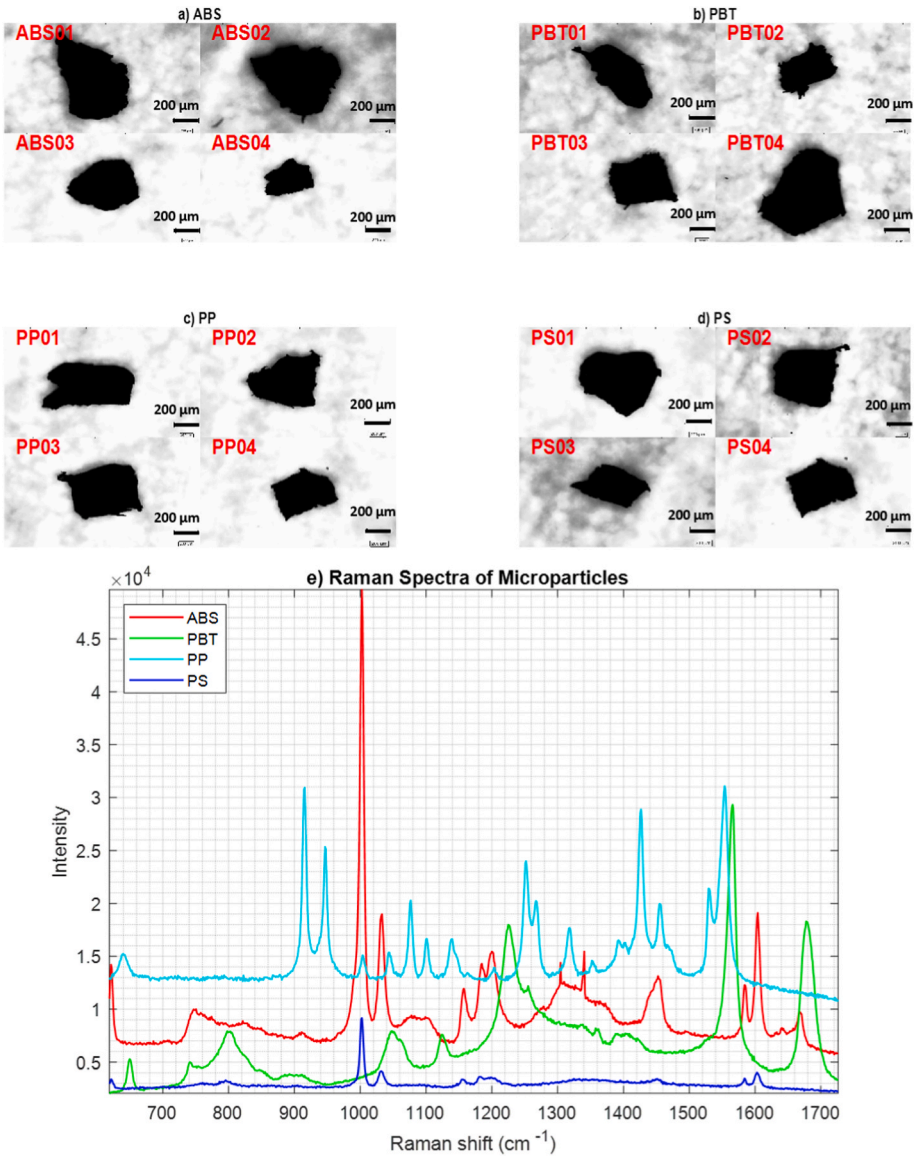

Fig. 1. a), b), c) and d) Microscopic still images of studied MPs. The scale bar is 200 µm for reference. e) Raman spectra of MPs of each polymer type confirms the identity of each sample, according to the peaks observed [42].

Table 1
Mean sizes and standard deviations of individual MPs seen in Fig. 1.

MP	Area ±standard deviation (μm^2)	Major Axis ±standard deviation (µm)	MP	Area ±standard deviation (μm^2)	Major Axis ±standard deviation (µm)
ABS01	160923 ± 2354	869 ± 5	**PP01**	114618 ± 1623	754 ± 3
ABS02	178743 ± 2923	747 ± 6	**PP02**	71880 ± 1172	656 ± 4
ABS03	103505 ± 1398	577 ± 2	**PP03**	98031 ± 1233	648 ± 4
ABS04	37328 ± 571	392 ± 2	**PP04**	55005 ± 1053	536 ± 3
PBT01	67718 ± 1464	712 ± 7	**PS01**	90214 ± 828	604 ± 2
PBT02	76235 ± 1580	542 ± 4	**PS02**	118367 ± 2211	594 ± 6
PBT03	90098 ± 1774	542 ± 4	**PS03**	46213 ± 1525	509 ± 6
PBT04	177028 ± 2477	879 ± 5	**PS04**	56000 ± 862	470 ± 3

After measuring the Raman spectrum and microscopic image of each MP, they were carefully placed on the surface of the sand, such that each Petri dish contained sand and one MP (red circle in Fig. 2). Reference pellets of the four polymers were placed alongside the Petri dishes on the bench-top setup to provide reference spectra for the training sets of classification models.

2.3. *Near-infrared hyperspectral imaging*

NIR-HSI images were collected using a HySpex NIR hyperspectral camera (Model: SWIR 384 S/N 3129, Oslo, Norway) with a 930 - 2500 nm spectral range. The lens had a focal distance of 30 cm and the conveyor belt moved at a rate of 1.00 cm/s. The spatial resolution was set to 200 µm. The light source was aimed at the sample at a 45° angle. Calibration was performed by taking a 0% reflectance measurement with the objective closed and 100% reflectance measurement with a white reference (a Spectralon plate). Using these two references, hyperspectral images were converted to reflectance with the following **Equation (1)**:

Fig. 2. (Left) Petri dish of sand doped with one ABS MP, circled in red, along with reference plastic pellet (white pellet on the top). (Right) NIR-HSI hyperspectral camera.

$$R_{(i,\lambda)} = \frac{C_{(i,\lambda)} - D_{(i,\lambda)}}{W_{(i,\lambda)} - D_{(i,\lambda)}} \quad \text{Equation 1}$$

for each pixel, *i,* and at each wavelength, λ, where $R_{(i,\lambda)}$ is the calculated reflectance; $C_{(i,\lambda)}$ is the intensity measured by the camera; $D_{(i,\lambda)}$ is the dark reference measurement representing background noise; $W_{(i,\lambda)}$ is the maximum reflectance intensity based on the measurement of the Spectralon plate. Each of these intensity values was collected for pixel i at λ wavelengths [43].

2.4. Calibration sets

As previously mentioned, each hyperspectral image was measured with a pellet of the corresponding pure polymer next to the Petri dish (Fig. 2) in order to use their spectra as a calibration set and avoid problems with variability (light conditions, focal distance or spatial accuracy of the camera). To create the training sets, around 100 spectra from sand and 50 spectra from the reference pellet of each polymer type were extracted to form a final spectral library of 150 spectra. All models were validated using random-subsets cross-validation and external validation (75% of the spectra for training and cross-validation, and 25% for external validation). The datasets for training and external validation were randomly split.

As sand is a heterogeneous multicomponent solid matrix, the selected spectra are not necessarily representative. Nevertheless, no trace organic matter remained due to the heat treatment performed in the sand. As such, any interference from the sand would likely arise from silicates, carbonates, or any of the other plausible components that might be found in a complex multicomponent matrix like sand.

Various pre-processing techniques were applied to the spectra (e.g. Standard Normal Variate, Extended Multiplicative Scatter correction, Savitzky-Golay first and second derivatives, and combinations of these). Because all models returned similar results, the simplest pre-processing method was selected—Standard Normal Variate. Mean centering was then applied to the pre-processed spectra prior to classification.

2.5. Classification models and classification assessment

2.5.1. Classification models

This manuscript compares the predictive ability of four well-known classification techniques: Soft Independent Modelling by Class Analogy (SIMCA), Partial Least Squares-Discriminant Analysis (PLS-DA), Support Vector Machines (SVM), and Artificial Neural Networks (ANN). The theoretical background of the classification models is well-known and has been widely published elsewhere [24,29,39,44–48]. The main details concerning their implementation and validation are provided here, but the interested reader is encouraged to consult the provided references for more detailed theoretical information.

The four algorithms were selected for their different approaches to developing classification models. Soft Independent Modelling by Class Analogy (SIMCA) [39] was performed by creating two independent classes for sand and the polymer. This class-modelling algorithm has the versatility of constructing independent sub-models for each class, exclusively using variance to construct boundaries between classes that allow projection of new samples in each sub-space. As such each sample is assessed for belonging to each class independently [30,48].

Unlike SIMCA, the other three classification methods are based on maximizing the differences between classes while minimizing the distances within each class. Partial Least Squares-Discriminant Analysis (PLS-DA) is the workhorse of linear classification methods [29,49,50]. Using PLS-DA, the properties of Linear Discriminant Analysis are combined with projection methods like Partial Least Squares. In this manuscript, PLS-DA models were optimized by selecting the lowest number of Latent Variables that gave the minimal classification error in external validation (test set).

Support Vector Machines (SVM) is often considered a non-linear algorithm [44,45]. Nevertheless, the classification is performed linearly in a sample hyperspace that has been rotated in a non-linear manner using kernel functions suited to specific problems. Two parameters are essential to optimize when choosing the support vectors: the margins between the two classes and the cost in the classification. The best model configuration obtained here was using PCA as data compression and Radial Basis Functions as the kernel option.

A classical but effective neural network approach has been used in this manuscript. Feed-forward backpropagation neural networks (ANN) are composed of an input, hidden, and output layer with the classification outcome [47,51]. The input layer contained the data compressed by a previous PCA model. The hidden layer was composed of two nodes. The transfer function between the nodes in the different layers was a sigmoidal logistic function.

2.5.2. Assessment of the classification power. Validation

All models were optimized using random sub-sets cross-validation and external test set validation. To construct the external validation set, 25% of the calibration data was left out. The validation parameters used in a classification model are Sensitivity, Specificity and Error Rate, all of

which are generally expressed in percentages. These parameters are based the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) during the internal cross-validation or external validation step [29]. Sensitivity refers to the ability of the model to confirm that a sample belongs to the class to which it has been assigned, and specificity is the ability of the model to confirm that a sample does not belong to the classes to which it has not been assigned [41].

One of the main features of classification models is their ability to calculate the probability of belonging to a given class for each pixel projected on the calibration model. For models like PLS-DA, SVM, and ANN, this probability is calculated in a binary problem as the distance of a specific sample to the center of a specific class divided by the sum of the distances of that sample to the center of the other classes:

$$P(A|y) = \frac{P(y|A)}{P(y|A) + P(y|B)} \quad \text{Equation 2}$$

This implies that the sum of the probabilities for both classes is equal to one (100%). Nevertheless, this is not the case for SIMCA, where independent class boundaries are created for each class and the probabilities are calculated independently for each class [48].

2.6. Data analysis and software

The hyperspectral images were pre-processed in MATLAB (MathWorks, Natick, MA, USA) using HYPER-Tools [52] (freely available at www.hypertools.org – last visited April 2023). SIMCA, PLS-DA, SVM, and ANN were applied using the PLS_Toolbox (Eigenvector Research, Inc., Manson, WA, USA).

3. Results and discussion

3.1. Spectral characteristics of polymers and sand

The NIR spectra of the pure plastics and a subset of spectra for the sand are shown in Fig. 3. Appreciable differences in the absorption bands and principal peaks are visible in the spectra, allowing these polymers to be differentiated. The spectral shape of ABS, PS, and PP can also be differentiated from the sand. ABS and PS share a number of absorption bands because styrene is a component of ABS.

It is important to highlight that PBT does not exhibit bands or spectral shapes that can aid in its identification, even though PBT has a well-known composition. The PBT pellets are black, and black plastics absorb NIR radiation at all of the wavelengths measured, making them notoriously difficult to measure.

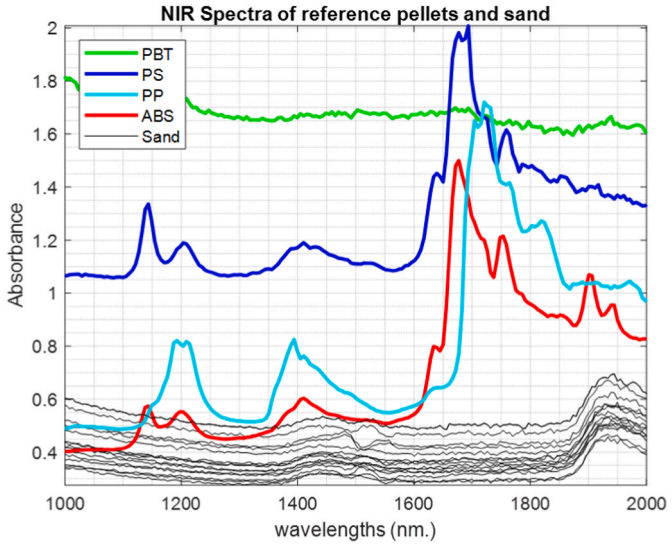

Fig. 3. Average NIR spectra of reference pellets of ABS, PBT, PP, and PS, as well as 15 spectra of sand from an area known not to contain plastic.

As previously mentioned, sand is a multicomponent system comprised of carbonates, silicates, and other inorganic matter. Fifteen randomly selected spectra of sand are shown in Fig. 3. All of the spectra look very similar, with only minor differences in some bands and the baseline drift characteristic of a scattering effect.

3.2. Training, cross-validation, and external validation

The results obtained for the calibration stage (CAL), together with cross-validation (CV) and external test set (TEST) indicated that perfect classification results, even in test sets, are achieved. Sensitivities and specificities of 100%, and classification errors of 0% were obtained for all the plastics and all the models tested. Despite this encouraging result, this might also indicate that the spectra used for the standard plastics span a very narrow variable space, since they all come from the same bulk of samples. This is a common practice in laboratories, since finding standard plastics of the same polymer with substantial differences in the spectra is a very difficult task.

In the next sections, the results for detecting ABS, PS, and PP will be carefully explained model by model (sections 3.3, 3.4, and 3.5), always referring to the supplied table and figures (in the manuscript and in the supplementary material). Nevertheless, the results for PBT with SIMCA (Fig. S09) and the rest of the model strategies will be treated in greater detail in section 3.6.

3.3. Prediction using SIMCA

Using SIMCA, objects—pixels, in this case—are assigned a probability of belonging to each of the classes in the model. The user may choose to look at the data through the lens of an individual class—e.g. the probability that a pixel contains the polymer of interest—or through the assignment of the most probable class, in which an object is assigned to the class it is most similar to, although objects may not be assigned to any class and remain "unassigned." Of the four classification models evaluated, SIMCA yielded the greatest number of true positive identifications of the MPs while minimizing the number of false positives. The SIMCA model located three of the four ABS MPs (Fig. 4, Fig. S01, Fig. S05 and Fig. S09).

We describe pixels as "positively identified" as belonging to a polymer class when there is a non-zero probability of class belonging, because the selection of thresholds is an open question. In this case, the threshold is the minimum probability of class belonging required to assign a pixel to a given polymer class. Setting too high of a threshold may cause pixels containing a MPs to be excluded and lead to underestimating the size. In contrast, setting the threshold too low may lead to false positives or overestimates of the size.

To further our discussion of probability thresholds, the size of MPs was calculated based on the pixels designated as having non-zero probability of containing polymers. In the cases of ABS01 and ABS02, the number of pixels characterized as ABS changes depending on where one sets the threshold for minimum probability of class belonging. ABS02, which has a major axis length of 747 ± 6 μm, will be used as an example. Five pixels with a non-zero probability of class belonging for the ABS class have been found (Fig. 4). The probabilities of class belonging are, in descending order: 0.4, 0.1, 0.08, 0.03, and 0.01. Each pixel is approximately 250 μm × 250 μm. All five pixels would be included if a threshold of 0.01 were applied, leading to a major axis length estimate of 750 μm. Vidal and Pasquini (2021) [21] applied a threshold of 0.01 for inclusion of pixels using a SIMCA model. In contrast, a threshold of 0.1 would include two pixels and exclude three pixels, leading to a major axis length estimate of 500 μm. Such decisions play an important role in estimating the size of MPs. In the case of ABS04, the SIMCA model located one pixel with a probability of class

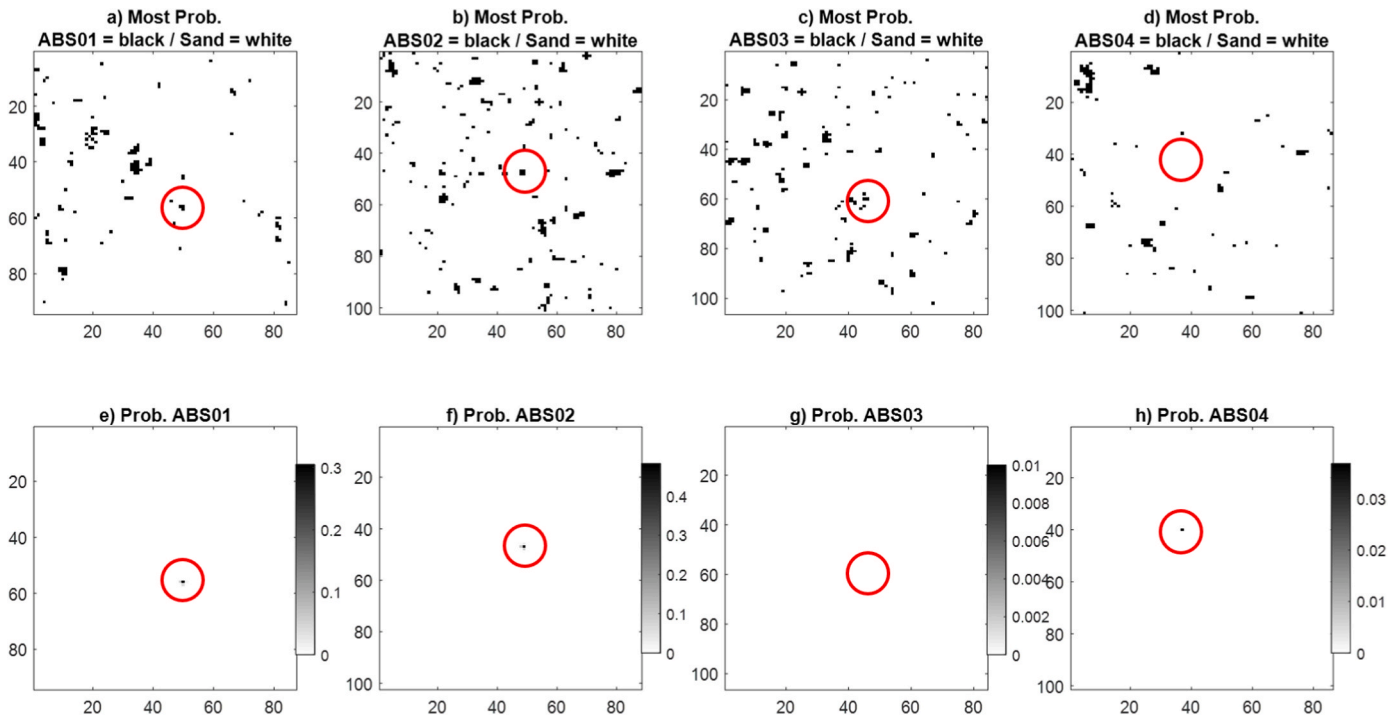

Fig. 4. SIMCA Most probable assignation (a, b, c, d) and Probability (e, f, g, h) for ABS.

belonging of 0.03 for ABS. In this case, the selection of a threshold for the minimum probability of class belonging can mark the difference between the measured presence and absence of MPs.

Although the major axis length and area of ABS03 were greater than those of ABS04 as measured via microscopy, SIMCA identified pixels of ABS04 but did not locate ABS03. Due to the surface roughness and the reflective nature of the sand, the ABS03 may have been oriented such that not all of the radiation that interacted with the MP was reflected to the camera sensor. This would be consistent with the light scattering and morphological effects that are known concerns in NIR-HSI. Similarly, PP02 was not located although it was not the smallest MP of PP by area or by major axis.

SIMCA located the MPs of PS in all four instances, but false positives were present in three of four cases (Fig. S01). For PS01, PS03, and PS04, multiple groups of pixels had non-zero probabilities of class belonging, indicating false positives. These pixels had probabilities of class belonging ranging from 0.04 to 0.6. In the case of PS03, there was one false positive, which was a pixel misclassified as PS with a probability of class belonging of 0.01. This further highlights the challenges of selecting a threshold for class inclusion and exclusion.

Regarding PP, the SIMCA model located the microplastics MPs in three of four PP cases: PP01, PP03, and PP04 (Fig. S05). There were no false positives. PP01, which has a major axis length of 754 ± 3 μm, is another example of a microplastic MP with a size estimate that depends on the selection of the minimum probability threshold for class belonging. SIMCA located five pixels with non-zero probability of class belonging for PP01 (Fig. S05): 0.7, 0.3, 0.1, 0.06, and 0.01. As such, a threshold of 0.1, for example, would lead to an estimate of major axis length of 500 μm instead of including all pixels, which would lead to an estimate of 750 μm. In the cases of PP03 and PP04, pixels were located with a non-zero chance of belonging to the PP class, although all had a probability of class belonging less than 0.03.

Using probabilities and selecting a threshold of class belonging is a normal practice when models like SIMCA are developed. Nevertheless, there is a probability for each pixel to belong to the category of sand, and special consideration should be given to the fact that even though the probabilities found for the pixels to belong to the polymer class (Fig. 4, Fig. S01 and Fig. S05), there might be the possibility of having a high probability of those pixels to be sand. This point is further discussed in section 3.7. The most probable plots clearly show that when the probability of belonging to the sand class is higher than the probability of belonging to the polymer class, a high number of false positives are found in all the cases studied (Fig. 4, Fig. S01 and Fig. S05).

3.4. Prediction using PLS-DA

The PLS-DA prediction model located three of the four ABS MPs (Fig. 5). However, the prediction model also yielded multiple false positives in all four cases.

In contrast, PLS-DA was only able to locate two of the four MPs of PS. False positives were observed in all the cases, even though there are fewer false positives than were obtained with SIMCA using most probable class assignation (Fig. S02). In the PP cases, PLS-DA was prone to false positives as well, although there were more false positives predicted with PLS-DA compared to SIMCA using most probable class assignation. (Fig. S06). The predictions erroneously identified various groups of pixels as corresponding to the class PP. Even so, PLS-DA did correctly locate the MPs of PP in all four instances.

3.5. Prediction using SVM

SVM was capable of correctly locating pixels with some probability of class belonging for all four MPs of each ABS, PP, and PS. Despite this initially encouraging outcome, there appear to be some drawbacks to the use of SVM prediction models. For example, more than 15 false positives (contiguous groups of pixels labeled as containing the polymer of interest) based on most probable class are observed in the images of all four MPs of ABS (Fig. 6).

The threshold selected for minimum probability of class belonging determines the number of MPs located by the SVM models. Deeper study is required to select a non-arbitrary threshold. To illustrate this point, the cases of PP, PS, and ABS will be compared. For PP and PS, the greatest probabilities of class belonging for any given pixel were 0.67 and 0.61, respectively (Figs. S03 and S07). If a 0.6 were selected as the

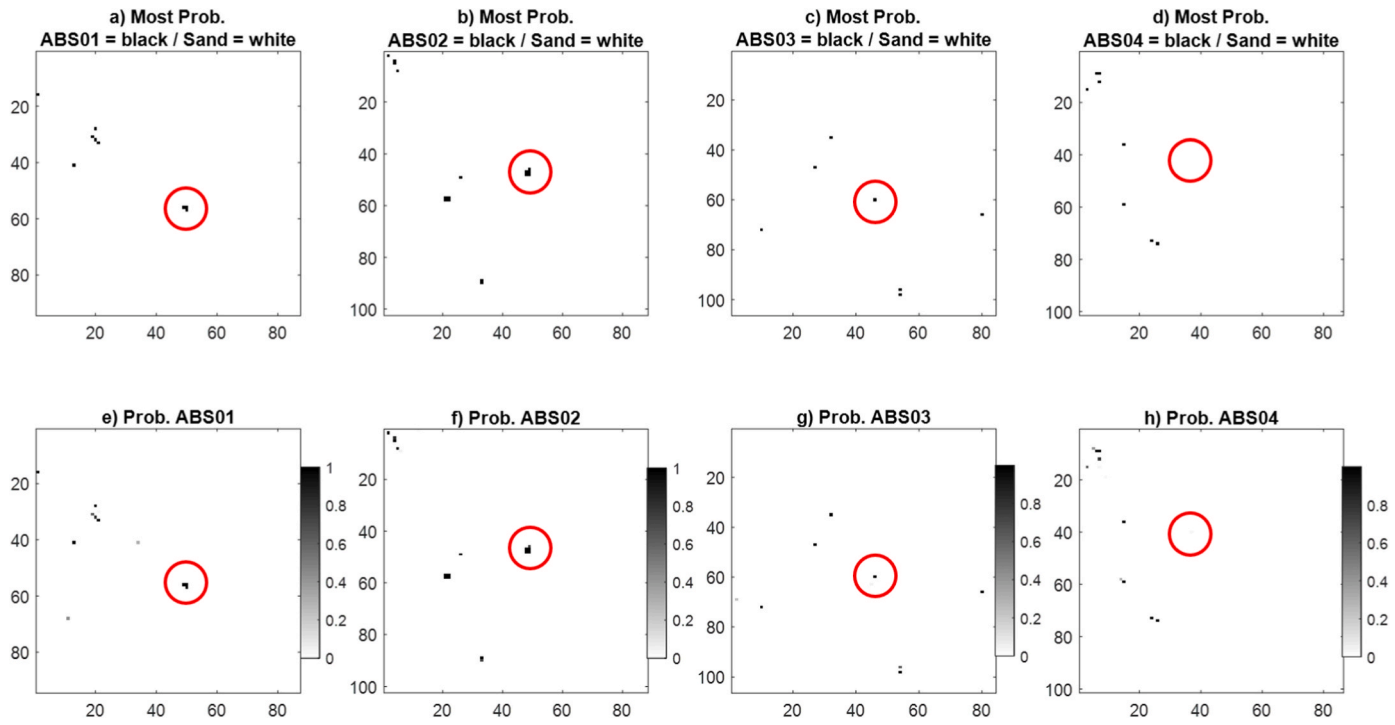

Fig. 5. PLS-DA Most probable assignation (a, b, c, d) and Probability (e, f, g, h) for ABS.

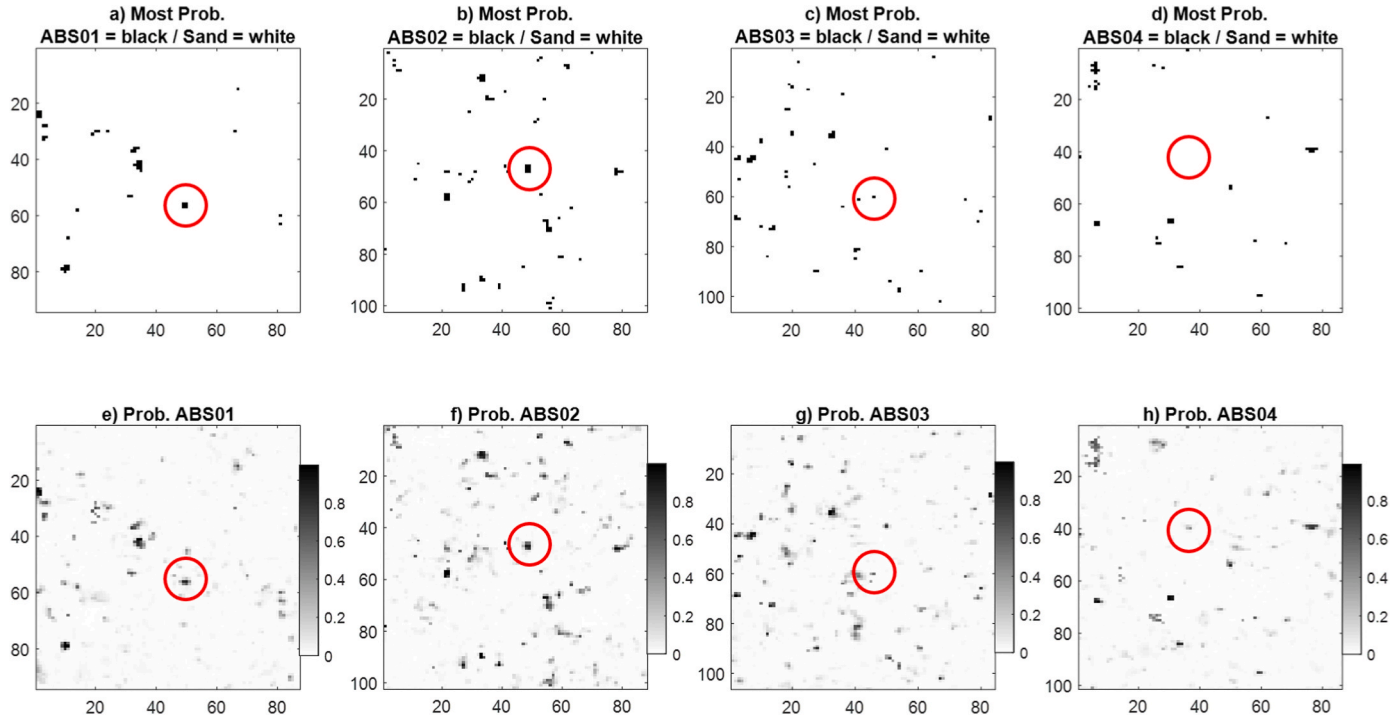

Fig. 6. SVM Most probable assignation (a, b, c, d) and Probability (e, f, g, h) for ABS.

minimum probability threshold, only one of four PP MPs and one of four PS MPs would be characterized as MPs. Meanwhile, using the same 0.6 threshold for ABS, upwards of ten groups of pixels would be labeled MPs, resulting in false positives (Fig. 6).

In all twelve plates of sand containing one MP each, SVM appears to classify the sand as having some probability of class belonging to the polymer classes. The SVM predictions appear to have a visual "texture" similar to that of sand, indicating that the models may be susceptible to mischaracterizing the scattering of the uneven surface of the sand.

3.6. *Prediction using ANN*

ANN models appear to fall short as a technique to locate MPs in sand. The ANN models correctly located two of the four ABS MPs (Fig. 7). However, in all four instances of ABS characterization, at least three false positives were identified as having non-zero probability of containing ABS. ANN correctly located two of the four MPs of ABS—ABS01 and ABS02 (Fig. 7). However, both particles are estimated to have a major axis length of two pixels, or approximately 500 μm. This is an

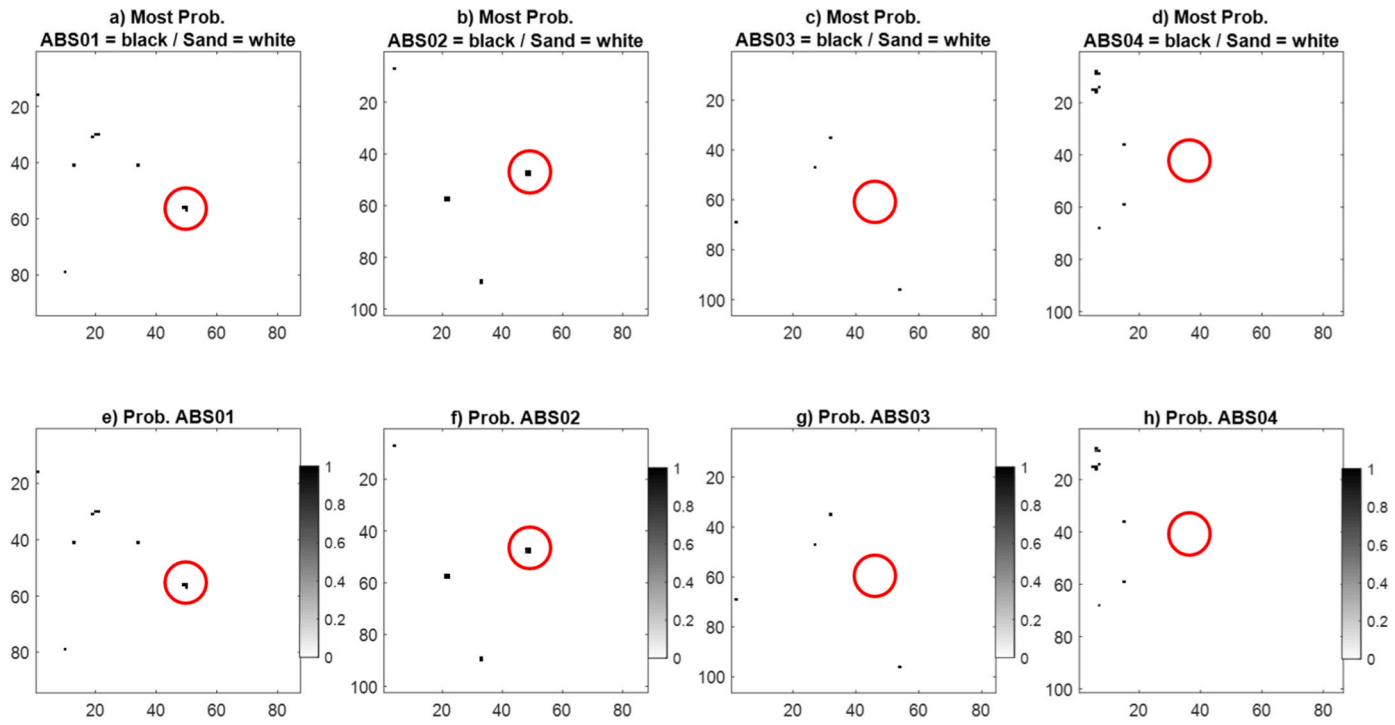

Fig. 7. ANN Most probable assignation (a, b, c, d) and Probability (e, f, g, h) for ABS.

underestimate, as ABS01 has a major axis length of 869 ± 5 μm and ABS02 has a major axis length of 747 ± 6 μm.

The ANN predictions for PS correctly identify the location of PS01 and PS02 with no false positives and the location of PS03 with one false positive (Fig. S04).

ANN locates one of the four PP MPs, PP01. No pixels are highlighted as containing PP for PP02 and PP04. In the case of PP03, ANN does not locate the MP. Instead, it locates a false positive. Curiously, the false positive is in the same location as the false positive for PP3 highlighted using SVM (Fig. S08).

3.7. Black PBT as a special case

As mentioned beforehand, there are no problems with the molecular structure of PBT (Fig. 1). Nevertheless, the main issue with this plastic is its black color. NIR spectroscopy is a photonic-based molecular technique. The black color causes most of the photons to be absorbed and transformed into energy, instead of being reflected. Therefore, the spectral signal is normally very poor quality. The results obtained by SIMCA (Fig. S09) show that the PBT MPs were not located under any circumstances. As such, whether or not PLS-DA, SVM or ANN correctly identified the location of the MP is unknown (Figure S10, S11 and S12). Regardless, it is clear that the models show optimistic results, as upwards of twenty connected groups of pixels were classified as PBT, even though only one particle of PBT was placed in the sand. The false positives may correspond to black or dark-colored grains of sand.

3.8. On spectra and probability

All cases presented in this manuscript (16 MPs of 4 polymers classified by 4 models) have presented, in general, difficulties in finding a unique classification of the MPs. Most cases, if not all, presented an elevated number of false positive pixels. One could ascribe this issue to the models themselves. Nevertheless, a deeper view of the spectra must be given in order to understand the root of the problem. NIR radiation working in diffuse reflectance may behave in a peculiar manner. It may penetrate thin MPs, be subject to influence from the scattering promoted by the background pixels, or even disappear in the shape of absorption as in the case of PBT. All these peculiarities make the spectra of the MP quite different from the ones in our databases, adversely affecting the likelihood of correct classification.

Let us take, as an example, the results obtained for the ABS01MP. Revisualizing the results obtained for that particular MP (Fig. 8), a closer detail can be given to the spectra associated with the pixels being classified as MP (red spectra in Fig. 8i) and the ones in the surroundings (green spectra in Fig. 8i). When compared with the standard reference spectra for ABS (thick red spectrum) and the mean spectrum of sand (thick green spectrum), it is evident that both groups of spectra share peaks belonging to ABS (e.g. the region between 1600 and 1900 nm) and baseline drifts belonging to the sand (e.g. the region between 1000 and 1200 nm). The models struggled to find those pixels whose spectra are a mixture of signals belonging to the plastic and the surroundings. For example, the blue spectrum belongs to the pixel classified as plastic by SVM only and not the other models. SVM found that the probability of that particular pixel to be either of the classes in 0.5, making its classification in one of the classes a mere question of mathematical computation precision rather than sound chemical criterion.

One might argue that feature selection could have helped improve the results in this case. Certainly, having all the available spectral information beforehand could help construct a more reliable model. Nevertheless, this is not possible for systems where the MPs are in environmental substrates like sand or even when the chemical structure of the polymer differs from the standard structures that we normally have in databases (presence of additives and colorants, action of weathering, biofouling, etc.). The presence or absence of different bands from the standard cannot be a priori predicted. Besides, it is important to remember that the classification results were of high quality (see Table 2). Therefore, there are no parameters left to optimize since the classification model is as optimal as possible as it is.

4. Conclusions

The detection of MPs in environmental substrates using chemical imaging and machine learning is increasing the interest of the scientific

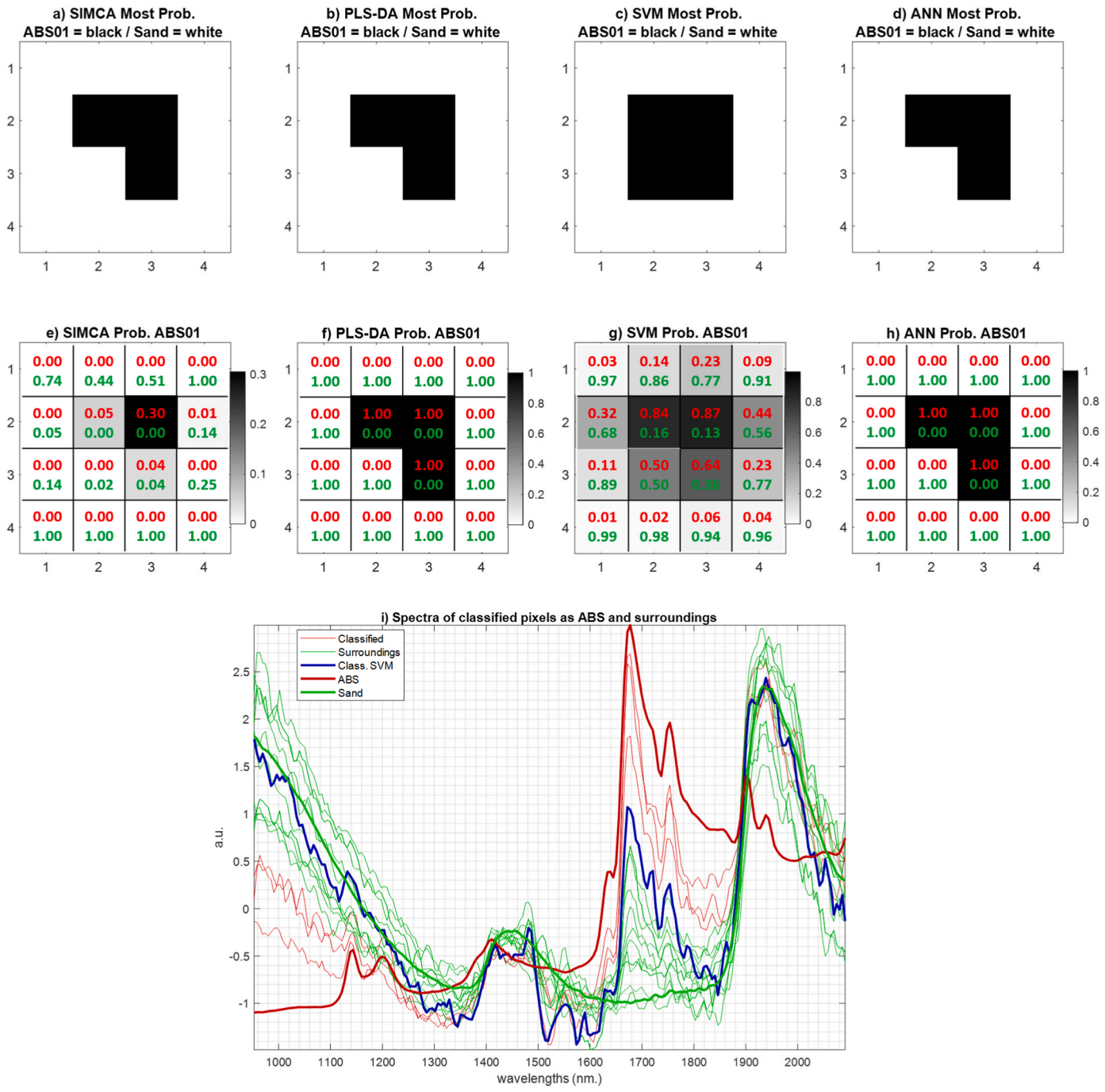

Fig. 8. A closer look at the results obtained for ABS01 with the four models. (a - d) the most probable class. (e - h) probability of belonging to the ABS class. In these cases, the actual probability calculated by the models for being ABS (red) and sand (green) is also displayed. i) Spectral profiles obtained for the pixels classified as ABS (red) and for the pixels surrounding them (green). The pixel classified as ABS by SVM that has not been classified as ABS by the other models has been drawn in blue. The pure spectrum for ABS is drawn with a thick red line, while an average spectrum for pure sand is drawn with a thick green line.

community. Nevertheless, it is not exempt from strong issues as demonstrated in this manuscript even with the simplest case that can be studied, since the MPs used here have the exact same composition as the pellets used for constructing the classification models. In this simple case, SIMCA was the most successful of the four machine learning techniques based on its ability to locate pixels containing the ABS, PS, and PP MPs while minimizing false positives, although some false positives were still observed. Within the discriminant analysis techniques, PLS-DA classification outperformed SVM and ANN, which both failed to locate the MPs in a greater number of cases. In the case of SVM, a notably higher number of false positives were observed.

The four classification techniques used are all supervised, pixel-based predictive models. SIMCA, a class-modelling technique, fundamentally differs from the other three classification techniques evaluated. Using SIMCA, pixels are evaluated for their similarity to a set number of principal components from a training set or spectral library on a pixel-by-pixel basis. This allows the construction of one-class models, and the addition of more classes without recalculating the entire model. In contrast, PLS-DA, SVM, and ANN are pure classification, or discriminant, techniques. With pure classification techniques, objects are sorted into classes based on the differences between classes present in the spectral library.

Table 2
Classification results. *For PBT, the results must be checked and understood with the supplied figures (S09 – S12) due to the extremely high number of false positives.

Particle	known MP in-class	SIMCA				PLS-DA				
		Area (μm^2) in-class object	Max Prob pixel*	Major Axis Length (μm)	False Positives?	known MP in-class	Area (μm^2) in-class object	Max Prob pixel*	Major Axis Length (μm)	False Positives?
ABS01	yes	250000	0.3	750	yes	yes	187500	1	500	yes
ABS02	yes	312500	0.4	750	yes	yes	312500	1	750	yes
ABS03	no	-	0	-	yes	yes	312500	0.8	750	yes
ABS04	no	-	0.03	-	yes	no	-	0	-	yes
PS01	yes	375000	0.5	750	yes	yes	250000	0.9	750	yes
PS02	yes	375000	0.6	750	no	yes	250000	1	500	yes
PS03	yes	62500	0.03	250	yes	no	-	0	-	yes
PS04	yes	125000	0.04	500	yes	no	-	0	-	yes
PP01	yes	312500	0.7	750	no	yes	437500	1	750	yes
PP02	no	-	0	-	no	yes	187500	1	500	yes
PP03	yes	125000	0.03	500	no	yes	125000	1	500	yes
PP04	yes	62500	0.01	250	no	yes	125000	1	500	yes
PBT01	yes*	-		-	yes	yes*	-		-	yes
PBT02	yes*	-		-	yes	yes*	-		-	yes
PBT03	yes*	-		-	yes	yes*	-		-	yes
PBT04	yes*	-		-	yes	yes*	-		-	yes
ABS01	yes	1000000	0.9	1000	yes	yes	187500	1	500	yes
ABS02	yes	1062500	0.8	1250	yes	yes	250000	1	500	yes
ABS03	yes	750000	0.6	1000	yes	no	-	0	-	yes
ABS04	no	-	0.4	-	yes	no	-	0	-	yes
PS01	no	-	0.4	-	yes	no	-	0	-	yes
PS02	yes	687500	0.6	250	yes	yes	250000	1	500	yes
PS03	no	-	0.2	-	yes	no	-	0	-	yes
PS04	no	-	0.4	-	yes	no	-	0	-	yes
PP01	yes	1187500	0.6	1250	yes	yes	125000	1	500	yes
PP02	no	-	0.3	-	yes	no	-	0	-	yes
PP03	no	-	0.4	-	yes	no	-	0	-	yes
PP04	no	-	0.25	-	yes	no	-	0	-	yes
PBT01	yes	187500	1	500	no	yes	187500	1	500	no
PBT02	yes	687500	0.9	250	no	yes	187500	1	500	no
PBT03	no	-	0.4	-	no	no	-	0	-	no
PBT04	no	-	0.4	-	no	no	-	0	-	no

Researchers must be aware of the potential pitfalls associated with classification based on libraries containing the spectra of pure polymers, which may create class boundaries too narrow to locate MPs effectively. This is particularly evident in this study, in which the MPs were not always located by the models, despite the fact that the MPs were chemically identical to the polymers in the spectral library. As such, weathering processes in the environment, chemical additives such as flame retardants, and the presence and formation of biofilm on plastics may be expected to negatively impact classification model performance. In the case of biofilms, the plastic signal is likely to be detectable because NIR radiation is known to penetrate samples. However, biofilm, like chemical additives, are likely to have NIR signal, in which case there would be a high likelihood of spectral interference that could impede correct classification. To address these concerns, other authors have noted the necessity of including weathered, commercial, primary, and secondary plastics with varying textures and densities in spectral libraries [25, 53], and our work supports this recommendation.

Also known to affect detection with NIR radiation are the spectral artifacts that arise from physical characteristics of MPs and sand. These physical influences include light scattering and morphological effects resulting from rough surfaces. In two instances, MPs were not located by SIMCA, despite not being the smallest MPs by area or diameter in each polymer class. As such, it is likely that the angle of incident radiation, the orientation of the MP, and the scattering from the rough surface of the sand may have interacted in such a way that the reflected radiation did not reach the camera sensor. Further, these results highlight the imprecise relationship between pixel size and detection limit for NIR-HSI. To this end, researchers must be careful when selecting thresholds of probability for class belonging. Default settings in classification tools allow the researcher to feel like they are avoiding making choices that may bias a model. However, not making a choice is also a choice, and may have ramifications such as over- or underestimates.

The combination of adequate machine learning classification models and NIR-HSI applied under appropriate circumstances has the potential to be a valuable addition to microplastic monitoring protocols, given its adaptability, potential to collect data for large areas in a short time, and low sample preparation requirements. However, for this to be the case, considerations about spectral libraries, spectral artifacts arising from physical characteristics, and threshold selection must be taken into account.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors did NOT use any Generative AI or AI-assisted technology in the writing process.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jose Manuel Amigo reports financial support was provided by Basque Government. Reaha Goyetche reports financial support was provided by University of the Basque Country. Leire Kortazar reports financial support was provided by Spain Ministry of Science and Innovation.

Data availability

Data will be made available on request.

Acknowledgements

This work was partially funded by Basque Government (KK 2021/00001 ELKARTEK 2021/2022). Reaha Goyetche thanks the University of the Basque Country, Spain, for her FPI grant. Leire Kortazar thanks the Spanish Ministry of Science and Innovation through project PID2020-118685RB-I00.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trac.2023.117221.

References

[1] C. Arthur, J. Baker, H. Bamford (Eds.), Proceedings of the International Research Workshop on Microplastic Marine Debris | OR&R's Marine Debris Program, 2008, in: https://marinedebris.noaa.gov/proceedings-international-research-workshop-microplastic-marine-debris. (Accessed 30 September 2021), in: https://marinedebris.noaa.gov/proceedings-international-research-workshop-microplastic-marine-debris.

[2] Sources GESAMP, Fate and Effects of Microplastics in the Marine Environment (Part 1), 2015. http://www.gesamp.org/publications/reports-and-studies-no-90. (Accessed 26 July 2021).

[3] K. Ashton, L. Holmes, A. Turner, Association of metals with plastic production pellets in the marine environment, Mar. Pollut. Bull. 60 (2010) 2050, https://doi.org/10.1016/j.marpolbul.2010.07.014. –2055.

[4] Y. Mato, T. Isobe, H. Takada, H. Kanehiro, C. Ohtake, T. Kaminuma, Plastic resin pellets as a transport medium for toxic chemicals in the marine environment, Environ. Sci. Technol. 35 (2001) 318–324, https://doi.org/10.1021/es0010498.

[5] L.M. Rios, C. Moore, P.R. Jones, Persistent organic pollutants carried by synthetic polymers in the ocean environment, Mar. Pollut. Bull. 54 (2007) 1230–1237, https://doi.org/10.1016/j.marpolbul.2007.03.022.

[6] S. Xu, J. Ma, R. Ji, K. Pan, A.-J. Miao, Microplastics in aquatic environments: occurrence, accumulation, and biological effects, Sci. Total Environ. 703 (2020), 134699, https://doi.org/10.1016/j.scitotenv.2019.134699.

[7] A.A. Horton, A. Walton, D.J. Spurgeon, E. Lahive, C. Svendsen, Microplastics in freshwater and terrestrial environments: evaluating the current understanding to identify the knowledge gaps and future research priorities, Sci. Total Environ. 586 (2017) 127–141, https://doi.org/10.1016/j.scitotenv.2017.01.190.

[8] K. Zhao, Y. Wei, J. Dong, P. Zhao, Y. Wang, X. Pan, J. Wang, Separation and characterization of microplastic and nanoplastic particles in marine environment, Environ. Pollut. 297 (2022), 118773, https://doi.org/10.1016/j.envpol.2021.118773.

[9] J. Frias, E. Pagter, R. Nash, I. O'Connor, O. Carretero, A. Filgueiras, L. Viñas, J. Gago, J. Antunes, F. Bessa, P. Sobral, A. Goruppi, V. Tirelli, M.L. Pedrotti, G. Suaria, S. Aliani, C. Lopes, J. Raimundo, M. Caetano, G. Gerdts, Standardised protocol for monitoring microplastics in sediments. https://doi.org/10.13140/RG.2.2.36256.89601/1, 2018.

[10] F. Ruggero, R. Gori, C. Lubello, Methodologies for microplastics recovery and identification in heterogeneous solid matrices: a review, J. Polym. Environ. 28 (2020) 739–748, https://doi.org/10.1007/s10924-019-01644-3.

[11] R. Mansa, S. Zou, Thermogravimetric analysis of microplastics: a mini review, Environ. Adv. 5 (2021), 100117, https://doi.org/10.1016/j.envadv.2021.100117.

[12] E.C. Minor, R. Lin, A. Burrows, E.M. Cooney, S. Grosshuesch, B. Lafrancois, An analysis of microlitter and microplastics from Lake Superior beach sand and surface-water, Sci. Total Environ. 744 (2020), 140824, https://doi.org/10.1016/j.scitotenv.2020.140824.

[13] L. Yang, Y. Zhang, S. Kang, Z. Wang, C. Wu, Microplastics in soil: a review on methods, occurrence, sources, and potential risk, Sci. Total Environ. 780 (2021), 146546, https://doi.org/10.1016/j.scitotenv.2021.146546.

[14] T. de J. Piñon-Colin, R. Rodriguez-Jimenez, M.A. Pastrana-Corral, E. Rogel-Hernandez, F.T. Wakida, Microplastics on sandy beaches of the baja California peninsula, Mexico, Mar. Pollut. Bull. 131 (2018) 63–71, https://doi.org/10.1016/j.marpolbul.2018.03.055.

[15] L. Cabernard, L. Roscher, C. Lorenz, G. Gerdts, S. Primpke, Comparison of Raman and fourier transform infrared spectroscopy for the quantification of microplastics in the aquatic environment, Environ. Sci. Technol. 52 (2018) 13279–13288, https://doi.org/10.1021/acs.est.8b03438.

[16] J. Workman, Interpretive spectroscopy of near infrared, Appl. Spectrosc. Rev. 31 (1996) 251–320, https://doi.org/10.1080/05704929608000571.

[17] F. Marini, J.M. Amigo, Chapter 2.4 - unsupervised exploration of hyperspectral and multispectral images, in: J.M. Amigo (Ed.), Data Handl. Sci. Technol, Elsevier, 2019, pp. 93–114, https://doi.org/10.1016/B978-0-444-63977-6.00006-7.

[18] M. Vidal, A. Gowen, J.M. Amigo, NIR hyperspectral imaging for plastics classification, NIR News 23 (2012) 13–15, https://doi.org/10.1255/nirn.1285.

[19] S. Piarulli, C. Malegori, F. Grasselli, L. Airoldi, S. Prati, R. Mazzeo, G. Sciutto, P. Oliveri, An effective strategy for the monitoring of microplastics in complex aquatic matrices: exploiting the potential of near infrared hyperspectral imaging (NIR-HSI), Chemosphere 286 (2022), 131861, https://doi.org/10.1016/j.chemosphere.2021.131861.

[20] S. Serranti, R. Palmieri, G. Bonifazi, A. Cózar, Characterization of microplastic litter from oceans by an innovative approach based on hyperspectral imaging, Waste Manag. 76 (2018) 117–125, https://doi.org/10.1016/j.wasman.2018.03.003.

[21] C. Vidal, C. Pasquini, A comprehensive and fast microplastics identification based on near-infrared hyperspectral imaging (HSI-NIR) and chemometrics, Environ. Pollut. 285 (2021), 117251, https://doi.org/10.1016/j.envpol.2021.117251.

[22] C. Zhu, Y. Kanaya, R. Nakajima, M. Tsuchiya, H. Nomaki, T. Kitahashi, K. Fujikura, Characterization of microplastics on filter substrates based on hyperspectral imaging: laboratory assessments, Environ. Pollut. 263 (2020), 114296, https://doi.org/10.1016/j.envpol.2020.114296.

[23] A. Hueni, S. Bertschi, Detection of sub-pixel plastic abundance on water surfaces using airborne imaging spectroscopy, in: IGARSS 2020 - 2020 IEEE Int. Geosci, Remote Sens. Symp., 2020, pp. 6325–6328, https://doi.org/10.1109/IGARSS39084.2020.9323556.

[24] V.H. da Silva, F. Murphy, J.M. Amigo, C. Stedmon, J. Strand, Classification and quantification of microplastics (<100 μm) using a focal plane array–fourier transform infrared imaging system and machine learning, Anal. Chem. 92 (2020) 13724–13733, https://doi.org/10.1021/acs.analchem.0c01324.

[25] L. Biermann, D. Clewley, V. Martinez-Vicente, K. Topouzelis, Finding plastic patches in coastal waters using optical satellite data, Sci. Rep. 10 (2020) 5364, https://doi.org/10.1038/s41598-020-62298-z.

[26] O. Garcia-Garin, T. Monleón-Getino, P. López-Brosa, A. Borrell, A. Aguilar, R. Borja-Robalino, L. Cardona, M. Vighi, Automatic detection and quantification of floating marine macro-litter in aerial images: introducing a novel deep learning approach connected to a web application in R, Environ. Pollut. 273 (2021), 116490, https://doi.org/10.1016/j.envpol.2021.116490.

[27] B.K. Veettil, N. Hong Quan, L.T. Hauser, D. Doan Van, N.X. Quang, Coastal and marine plastic litter monitoring using remote sensing: a review, Estuar. Coast Shelf Sci. 279 (2022), 108160, https://doi.org/10.1016/j.ecss.2022.108160.

[28] J.M. Amigo, Chapter 1.1 - hyperspectral and multispectral imaging: setting the scene, in: J.M. Amigo (Ed.), Data Handl. Sci. Technol, Elsevier, 2020, pp. 3–16, https://doi.org/10.1016/B978-0-444-63977-6.00001-8.

[29] D. Ballabio, V. Consonni, Classification tools in chemistry. Part 1: linear models. PLS-DA, Anal. Methods 5 (2013) 3790, https://doi.org/10.1039/c3ay40582f.

[30] F. Marini, Classification methods in chemometrics, Curr. Anal. Chem. 6 (2010) 72–79.

[31] A. Ulrici, S. Serranti, C. Ferrari, D. Cesare, G. Foca, G. Bonifazi, Efficient chemometric strategies for PET–PLA discrimination in recycling plants using hyperspectral imaging, Chemometr. Intell. Lab. Syst. 122 (2013) 31–39, https://doi.org/10.1016/j.chemolab.2013.01.001.

[32] S. Freitas, H. Silva, E. Silva, Remote hyperspectral imaging acquisition and characterization for marine litter detection, Rem. Sens. 13 (2021) 2536, https://doi.org/10.3390/rs13132536.

[33] A.P.M. Michel, A.E. Morrison, V.L. Preston, C.T. Marx, B.C. Colson, H.K. White, Rapid identification of marine plastic debris via spectroscopic techniques and machine learning classifiers, Environ. Sci. Technol. 54 (2020) 10630–10637, https://doi.org/10.1021/acs.est.0c02099.

[34] Y. Zhang, X. Wang, J. Shan, J. Zhao, W. Zhang, L. Liu, F. Wu, Hyperspectral imaging based method for rapid detection of microplastics in the intestinal tracts of fish, Environ. Sci. Technol. 53 (2019) 5151–5158, https://doi.org/10.1021/acs.est.8b07321.

[35] W. Ng, B. Minasny, A. McBratney, Convolutional neural network for soil microplastic contamination screening using infrared spectroscopy, Sci. Total Environ. 702 (2020), 134723, https://doi.org/10.1016/j.scitotenv.2019.134723.

[36] W. Ai, G. Chen, X. Yue, J. Wang, Application of hyperspectral and deep learning in farmland soil microplastic detection, J. Hazard Mater. 445 (2023), 130568, https://doi.org/10.1016/j.jhazmat.2022.130568.

[37] W. Ai, S. Liu, H. Liao, J. Du, Y. Cai, C. Liao, H. Shi, Y. Lin, M. Junaid, X. Yue, J. Wang, Application of hyperspectral imaging technology in the rapid

identification of microplastics in farmland soil, Sci. Total Environ. 807 (2022), 151030, https://doi.org/10.1016/j.scitotenv.2021.151030.
[38] A.C. Olivieri, G.M. Escandar, Chapter 6 - analytical figures of merit, in: A. C. Olivieri, G.M. Escandar (Eds.), Pract. Three-Way Calibration, Elsevier, Boston, 2014, pp. 93–107, https://doi.org/10.1016/B978-0-12-410408-2.00006-5.
[39] S. Wold, M. Sjöström, SIMCA: a method for analyzing chemical data in terms of similarity and analogy, in: Chemom. Theory Appl, American Chemical Society, 1977, pp. 243–282, https://doi.org/10.1021/bk-1977-0052.ch012.
[40] PlasticsEurope, Plastics - the Facts 2020: an Analysis of European Plastics Production, Demand and Waste Data, PlasticsEurope Association of Plastic Manufacturers, Brussels, Belgium, 2020. https://issuu.com/plasticseuropeebook/docs/plastics_the_facts-web-dec2020. (Accessed 31 January 2022).
[41] J.M. Amigo, H. Babamoradi, S. Elcoroaristizabal, Hyperspectral image analysis. A tutorial, Anal. Chim. Acta 896 (2015) 34–51, https://doi.org/10.1016/j.aca.2015.09.030.
[42] W. Cowger, Z. Steinmetz, A. Gray, K. Munno, J. Lynch, H. Hapich, S. Primpke, H. De Frond, C. Rochman, O. Herodotou, Microplastic spectral classification needs an open source community: open specy to the rescue, Anal. Chem. 93 (2021) 7543–7548, https://doi.org/10.1021/acs.analchem.1c00123.
[43] J. Burger, P. Geladi, Hyperspectral NIR image regression part I: calibration and correction, J. Chemom. 19 (2005) 355–363, https://doi.org/10.1002/cem.938.
[44] V. Vapnik, R. Izmailov, Reinforced SVM method and memorization mechanisms, Pattern Recogn. 119 (2021), 108018, https://doi.org/10.1016/j.patcog.2021.108018.
[45] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297, https://doi.org/10.1023/A:1022627411411.
[46] M.I. Jordan, T.M. Mitchell, Machine learning: trends, perspectives, and prospects, Science 349 (2015) 255–260, https://doi.org/10.1126/science.aaa8415.
[47] M.A. Mazurowski, P.A. Habas, J.M. Zurada, J.Y. Lo, J.A. Baker, G.D. Tourassi, Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance, Neural Netw. Off. J. Int. Neural Netw. Soc. 21 (2008) 427–436, https://doi.org/10.1016/j.neunet.2007.12.031.
[48] R. Vitale, M. Cocchi, A. Biancolillo, C. Ruckebusch, F. Marini, Class modelling by soft independent modelling of class analogy: why, when, how? A tutorial, Anal. Chim. Acta (2023), 341304, https://doi.org/10.1016/j.aca.2023.341304.
[49] D. Ruiz-Perez, H. Guan, P. Madhivanan, K. Mathee, G. Narasimhan, So you think you can PLS-DA? BMC Bioinf. 21 (2020) 2, https://doi.org/10.1186/s12859-019-3310-7.
[50] J. Gottfries, K. Blennow, A. Wallin, C.G. Gottfries, Diagnosis of dementias using partial least squares discriminant analysis, Dement. Basel Switz. 6 (1995) 83–88, https://doi.org/10.1159/000106926.
[51] J. Sietsma, R.J.F. Dow, Creating artificial neural networks that generalize, Neural Network. 4 (1991) 67–79, https://doi.org/10.1016/0893-6080(91)90033-2.
[52] N. Mobaraki, J.M. Amigo, HYPER-Tools. A graphical user-friendly interface for hyperspectral image analysis, Chemometr. Intell. Lab. Syst. 172 (2018) 174–187, https://doi.org/10.1016/j.chemolab.2017.11.003.