Peer Reviewed Paper Highlighting Young Investigators openaccess

Development of a classification algorithm for efficient handling of multiple classes in sorting systems based on hyperspectral imaging

Rosalba Calvini,[a,*] Giorgia Orlandi,[b] Giorgia Foca[c] and Alessandro Ulrici[d]

Department of Life Sciences, University of Modena and Reggio Emilia, Padiglione Besta, Via Amendola, 2, 42122 Reggio Emilia, Italy.
E-mail: rosalba.calvini@unimore.it
[a]: https://orcid.org/0000-0002-0100-8798, [b]: https://orcid.org/0000-0002-9116-4566
[c]: https://orcid.org/0000-0002-4436-774X, [d]: https://orcid.org/0000-0002-6010-6400

When dealing with practical applications of hyperspectral imaging, the development of efficient, fast and flexible classification algorithms is of the utmost importance. Indeed, the optimal classification method should be able, in a reasonable time, to maximise the separation between the classes of interest and, at the same time, to correctly reject possible outlier samples. To this aim, a new extension of Partial Least Squares Discriminant Analysis (PLS-DA), namely Soft PLS-DA, has been implemented. The basic engine of Soft PLS-DA is the same as PLS-DA, but class assignment is subjected to some additional criteria which allow samples not belonging to the target classes to be identified and rejected. The proposed approach was tested on a real case study of plastic waste sorting based on near infrared hyperspectral imaging. Household plastic waste objects made of the six recyclable plastic polymers commonly used for packaging were collected and imaged using a hyperspectral camera mounted on an industrial sorting system. In addition, paper and not recyclable plastics were also considered as potential foreign materials that are commonly found in plastic waste. For classification purposes, the Soft PLS-DA algorithm was integrated into a hierarchical classification tree for the discrimination of the different plastic polymers. Furthermore, Soft PLS-DA was also coupled with sparse-based variable selection to identify the relevant variables involved in the classification and to speed up the sorting process. The tree-structured classification model was successfully validated both on a test set of representative spectra of each material for a quantitative evaluation, and at the pixel level on a set of hyperspectral images for a qualitative assessment.

Keywords: PLS-DA, multivariate classification, hierarchical classification, sparse methods, feature selection, plastic sorting

Introduction

Over the past decades, Hyperspectral Imaging (HSI) has gained increasing attention from industries interested in the implementation of automated sorting systems to solve a number of different problems. Indeed, HSI has found a wide range of applications in the food industry, including the quality evaluation and safety assessment of several food products,[1] such as fruits and vegetables,[2,3] meat,[4] cereals[5] and dairy products.[6] Moreover, other manufacturing environments, such as the pharmaceutical industry, have employed real-time HSI systems for quality control and process monitoring in the frame of the process analytical technology.[7–9] Another relevant

Correspondence
Rosalba Calvini (rosalba.calvini@unimore.it)

Received: 8 November 2018
Revised: 10 December 2018
Accepted: 10 December 2018
Publication: 18 December 2018
doi: 10.1255/jsi.2018.a13
ISSN: 2040-4565

Citation
R. Calvini, G. Orlandi, G. Foca and A. Ulrici, "Development of a classification algorithm for efficient handling of multiple classes in sorting systems based on hyperspectral imaging", *J. Spectral Imaging* **7,** a13 (2018). https://doi.org/10.1255/jsi.2018.a13

field of application of HSI is represented by the recycling industry, where hyperspectral sensors are used to separate end-of-life objects, such as plastic,[10] paper[11] or electronic waste,[12] according to material type.

In these contexts, HSI can be considered as a step forward with respect to traditional spectroscopic techniques, which allow fast and non-destructive characterisation of the chemical properties of the analysed samples. In fact, HSI systems couple these advantages with the possibility of also visualising the spatial distribution of the chemical features of interest within the sample surface. Furthermore, in sorting systems, HSI can also be employed to quickly identify the chemical composition of homogeneous objects moving on a conveyor belt, and to distinguish them from samples with different composition.[13]

In practical situations, hyperspectral imaging can be applied to address complex classification issues, where the sorting problem under investigation requires the discrimination of several classes at the same time, with some classes sharing similar features. This can be easily managed by using HSI systems, since with a single measurement, i.e. with the acquisition of a single hyperspectral image, it is possible to have a wide range of information. However, in order to meet the needs of real-time applications, it is necessary to identify classification strategies able to handle a huge amount of spectral data, providing reliable results in short computational times.[14]

When dealing with multiple classes, this issue can be addressed using a tree-structured classification model, where each branching (tree node) corresponds to a local classification model.

In this manner, classification is performed considering a top-down approach, where the samples are initially assigned to general macro-categories, and then each macro-class is split into increasingly specific categories, until reaching the classes of interest.[15–17]

Another relevant issue to be faced in practical applications of HSI in sorting systems is related to the fact that, generally, it is not easy to have a strict control of the input stream in order to avoid the presence of foreign objects, i.e. objects not belonging to the target classes of the specific application. In this context, the availability of algorithms able to maximise the discrimination between the categories of interest and, at the same time, to identify possible foreign materials is of the utmost importance.

Partial Least Squares Discriminant Analysis (PLS-DA) is one of the most widely used methods for multivariate classification of hyperspectral data.[18] Basically, PLS-DA is an extension of the PLS algorithm, which aims at identifying a new set of variables, named Latent Variables (LVs), by maximising the between-classes variance. Class membership is coded using a dummy **Y** matrix, and the assignment of unknown samples is based on the *a posteriori* probability associated with the corresponding **Y** predicted values. The standard PLS-DA approach assigns a sample to the class for which it has the higher *a posteriori* probability, resulting in unknown samples always being assigned to one of the target classes.[19,20] Therefore, on the one hand PLS-DA has the great advantage of maximising the separation between the considered classes but, on the other hand, it does not allow a straightforward identification of outliers.

Conversely, the possibility of having unassigned samples is one of the major advantages of the so-called class-modelling techniques, which are essentially based on describing each single class independently from the others, and then verifying whether an unknown sample is compliant or not with the characteristics of each class of interest.[21] In this manner, it is possible that a new unknown sample is rejected from all the class models, resulting in an unassigned sample. Soft Independent Modelling of Class Analogy (SIMCA) is the most common class-modelling method. It calculates local Principal Component Analysis (PCA) models for each considered class, which are used to define class boundaries based on the distances both in the score space (Hotelling's T^2) and in the residual space (Q residuals).[22] Notwithstanding the advantages of class-modelling methods like SIMCA, they can provide poor classification results when the modelled classes are quite overlapped, since the model is not oriented towards the discrimination of the considered categories.

Given these considerations, it is reasonable to assume that a classification algorithm to be efficiently employed in sorting systems should comprise the advantages of both classification techniques and of class-modelling methods, i.e. it should be able to maximise the discrimination between the categories of interest and to recognise and reject outlier samples at the same time.

To this aim, in the present paper a modified version of the PLS-DA algorithm, namely Soft PLS-DA, is proposed. The basic principle of Soft PLS-DA is the same as PLS-DA, but class assignment is performed by fixing additional limits both on the **Y** predicted values and on the Q residuals. In this manner, the classification model is built

by maximising the differences between the modelled classes; at the same time, the additional limits allow the rejection of samples belonging to unexpected categories and relegation of them to a general category of unassigned samples.

The effectiveness of Soft PLS-DA algorithm was tested on a case study related to the implementation of a near infrared (NIR) hyperspectral imaging system for plastic waste sorting. Indeed, the different plastic polymers have a specific spectral fingerprint in the NIR range and optical sorting is commonly used to separate them.[23,24] The goal of our study consisted in the implementation of a classification method able to effectively discriminate paper and six recyclable plastic polymers commonly used for packaging, and to correctly reject objects belonging to non-target classes, such as non-recyclable plastics.

The present manuscript is structured as follows:

- The next section reports the theoretical background of the standard classification approach based on PLS-DA and a detailed description of the novel Soft PLS-DA algorithm.
- Material and methods describes the plastic dataset, the procedure followed for image acquisition and elaboration together with the different steps of data analysis.
- Results shows the classification results obtained using Soft PLS-DA algorithm both considering the full wavelength range and sparse-based variable selection, and also the results of the final implementation of the classification tree.
- Finally, we report the general conclusions of this study.

Theory

Background

PLS-DA is as an extension of PLS regression adapted to operate in a classification framework. Similarly to PLS, the independent matrix $\mathbf{X}$ is regressed against the dependent matrix $\mathbf{Y}$ by calculating a new set of LVs, which maximise the covariance between $\mathbf{X}$ and $\mathbf{Y}$ matrices.[25,26] In more detail, both $\mathbf{X}$ and $\mathbf{Y}$ matrices can be decomposed as follows:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{1}$$

where $\mathbf{T}$, $\mathbf{P}$ and $\mathbf{E}$ represent the score matrix, the loading matrix and the residuals matrix of $\mathbf{X}$, respectively;

$$\mathbf{Y} = \mathbf{U}\mathbf{C}^{\mathrm{T}} + \mathbf{G} \tag{2}$$

where $\mathbf{U}$ is the score matrix, $\mathbf{C}$ is the loading matrix and $\mathbf{G}$ is the residuals matrix referred to $\mathbf{Y}$.

According to PLS, the object variation of the $\mathbf{X}$-block expressed by the score matrix $\mathbf{T}$ can be used to describe $\mathbf{Y}$; therefore Equation 2 can be re-written as follows:

$$\mathbf{Y} = \mathbf{T}\mathbf{C}^{\mathrm{T}} + \mathbf{G} \tag{3}$$

In particular, the decomposition of $\mathbf{X}$ has to be optimised in a manner that $\mathbf{T}$ accounts for the variation in $\mathbf{X}$ which allows the best description of $\mathbf{Y}$. To this aim, for each LV a weight vector ($\mathbf{w}$) is calculated which weights the original variables according to their contribution in explaining the $\mathbf{Y}$ matrix. Given these considerations, the estimate of $\mathbf{Y}$ ($\hat{\mathbf{Y}}$) can be calculated as follows:

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{W}\mathbf{C}^{\mathrm{T}} + \mathbf{F} = \mathbf{X}\mathbf{B} + \mathbf{F} \tag{4}$$

where $\mathbf{W}$ is the weights matrix and $\mathbf{B}$ is the matrix of regression coefficients.

In the case of PLS-DA, the $\mathbf{Y}$ matrix consists of a dummy matrix with as many rows as the number of samples and as many columns as the number of considered classes. This dummy matrix expresses class membership of each sample with binary coding: a value equal to 1 indicates that an object belongs to the class, while a value equal to 0 refers to samples not belonging to the class.[19,27]

Once the PLS model has been calibrated, class assignment of unknown samples is based on the *y* values estimated for each class ($\hat{y}$). These values will not be exactly equal to 0 or 1, therefore it is necessary to establish a threshold value so that a new sample is assigned to a defined class only if its $\hat{y}$ value is greater than the threshold for that class. The threshold is usually calculated using the Bayes theorem under the assumption that the estimated values for each class follow a Gaussian distribution, and these distributions are used to calculate the *a posteriori* probability that a sample belongs to a given class.[20] In particular, for each class the threshold value corresponds to the $\hat{y}$ value at which the number of false positives and false negatives is minimised, that is the point where the two probability distributions cross (i.e. the point where the *a posteriori* probability values for the two classes are the same).

Usually, class assignment of an unknown sample can be done considering two different approaches: either choosing the class with the highest probability, or comparing the predicted $\hat{y}$ values with the corresponding threshold values. The former strategy represents the standard discriminant approach, where samples are always

assigned to one of the modelled classes. Conversely, if the latter approach is used, a sample can have $\hat{y}$ values higher than the corresponding threshold for more than one class, or lower than the threshold for all the classes. In both these cases, the sample cannot be assigned to any class.[27] However, in the case of only two classes, these two approaches converge to the same results and standard PLS-DA algorithm will always attribute a sample to one of the classes.

The standard discriminant approach has often been criticised due to its limited ability to correctly handle new objects not belonging to the target classes.[28] In order to overcome this issue, extensions of PLS-DA have been proposed in the literature, which incorporate a rejection option in the classification rule.

These methods usually consist in considering PLS-DA as a data compression method rather than a classification strategy, and class assignment is performed with a further class-modelling step considering distance-based metrics calculated on the PLS scores[16] or on the PCA scores obtained from the decomposition of the $\hat{\mathbf{Y}}$ matrix.[29] In this manner, class assignment is performed in a rather complex multi-step procedure.

Conversely, an alternative approach consists in calculating a classification rule considering confidence intervals around the $\hat{y}$ values of each class, and rejecting samples outside these intervals.[30]

Furthermore, it has to be considered that diagnostics based on Q residuals can represent an effective tool to identify outlier samples, i.e. samples with properties different from those of the samples used for model calibration. These diagnostics are widely used in class-modelling, for example in the SIMCA algorithm. However, Q residuals are rarely incorporated in classification rules based on PLS-DA algorithm, even if the computation of Q scores from the outcomes of the PLS model is straightforward according to the following equation:

$$Q_i = e_i e_i^T \tag{5}$$

where Q_i is the Q score value of the i^{th} sample of matrix $\mathbf{X}$ and e_i is the i^{th} row vector of the residuals matrix $\mathbf{E}$.

Soft PLS-DA

In the present study, a novel algorithm based on PLS-DA has been developed in order to combine the advantages of classical discriminant analysis with those of class-modelling techniques; for these reasons it has been named Soft PLS-DA. The main idea behind Soft PLS-DA is to have a flexible and simple classification tool able to maximise the separation between the considered classes and, at the same time, to effectively identity possible outlier objects (i.e. objects that do not belong to the classes included in the classification model), which will be automatically not assigned to any class.

In the same manner as PLS-DA, a PLS model is calculated between the $\mathbf{X}$ matrix and the dummy $\mathbf{Y}$ matrix, in order to maximise the differences between the considered classes. Then, class assignment of unknown samples is based on some additional criteria which allow outlier samples to not be assigned to the target classes.

According to the Soft PLS-DA decision rules, class assignment of a new sample to a defined class is subjected to the following criteria:

- having Q residuals falling inside the 99.9 % confidence limit of the model.[31] The 99.9 % confidence limit has been chosen in order to set boundaries large enough to consider as much as possible the variability of the different classes and, at the same time, being able to exclude samples with a very low fit to the model;
- having $\hat{y}$ values falling inside an acceptability range for the considered class. More in detail, in addition to the threshold value calculated by standard PLS-DA for each class g ($y_{tsh1,g}$), also an upper limit ($y_{tsh2,g}$) on the $\hat{y}$ values has been introduced, which is calculated as follows:

$$y_{tsh2,g} = m_{\hat{y},g} + 5 \times s_{\hat{y},g} \tag{6}$$

where $m_{y\hat{},g}$ and $s_{y\hat{},g}$ are the mean and the standard deviation of the $\hat{y}$ values of class g calculated on the training set samples. Therefore, in order to be assigned to class g, an unknown sample must have a $\hat{y}$ value ranging between $y_{tsh1,g}$ and $y_{tsh2,g}$. The upper limit imposed on the $\hat{y}$ values allows objects found at the extremes of the Gaussian probability density functions (PDFs) to be rejected; these usually have low values in the PDFs but high *a posteriori* probability for one class according to the Bayes rule. The upper limit was set based on preliminary tests performed on some representative images;

- for classification problems with more than two classes, the samples must be unambiguously assigned only to one class.

Samples that do not match all the three criteria are not assigned to any class and are labelled as "not assigned" (NA). In this manner, Soft PLS-DA allows boundaries to be drawn around each modelled class which maximise the discrimination between the categories of interest and

minimise possible false positives due to ambiguous classifications or to outlier samples.

Material and methods

Plastic dataset

In the present study, we have considered the recyclable plastic polymers mainly used for packaging, including polyethylene terephthalate (PET), polystyrene (PS), polyvinyl chloride (PVC), polypropylene (PP) and polyethylene (PE), which in turn can be further subdivided in high-density polyethylene (HDPE) and low-density polyethylene (LDPE).

Different plastic objects made of the considered polymers have been collected form household waste. In addition, samples composed of paper and of other types of non-recyclable plastics (OTHER), e.g. acrylonitrile butadiene styrene (ABS) and polylactic acid (PLA), were also considered as possible foreign materials that can be found in plastic municipal waste.

The different samples have been manually sorted into the corresponding categories based on the Resin Identification Code (RIC) reported on the objects.[32] The RIC is an international coding system comprising a set of symbols (labels and numbers) present on plastic products and indicating the polymer of which they are composed. For example, according to RIC, coding number 1 is associated to PET, number 2 is associated to HDPE, number 3 is associated to PVC etc.

Image acquisition

The collected waste samples have been acquired with an industrial sorting system consisting of a NIR line scanning hyperspectral camera (KUSTA1.9MSI, LLA Instruments) mounted over a black conveyor belt and equipped with an InGaAs detector array and Zeiss f/2.4, 10 mm optical lens. Image acquisition was performed with a frame rate equal to 644 Hz and the speed of the conveyor belt was equal to 0.84 m s^{-1}. Illumination was provided by halogen light bulbs positioned in two parallel illumination rows slightly tilted towards each other (PMAmsi, LLA Instruments). The hyperspectral images were acquired in the NIR range from 1330 nm to 1900 nm with a spectral resolution of 6 nm.

The samples were acquired in two acquisition phases conducted on different days. In the first phase, hyperspectral images containing objects made of the same material were acquired. In more detail, two images of different objects were acquired for each type of material, for a total of 16 hyperspectral images (= 8 materials × 2 replicated images). These images were used as training images in the subsequent elaboration steps, in order to obtain a library of representative spectra for each material type.

In the second acquisition phase, hyperspectral images containing samples of two different materials were acquired considering all the possible combinations between the material types under investigation. On the whole, 56 hyperspectral images have been obtained in this phase, resulting from two replicate images for each combination.

All the hyperspectral images, acquired on the objects positioned on the moving conveyor belt, have size equal to 41 row pixels × 500 column pixels × 96 wavelengths. For each image, the raw intensity counts were converted into reflectance units by means of an internal calibration procedure based on the measure of the dark current and of a white high reflectance standard. Dark current was measured by closing the shutter of the camera, while the white reference consisted of an aluminium frame holding the calibration material with an average remission factor equal to 83.0 %.

Image elaboration

Initially, the 16 images acquired during the first acquisition phase were analysed by means of PCA after mean centring as data preprocessing, in order to segment the pixels of the background (black conveyor belt) from those belonging to the samples.

After background segmentation, from each training image 1000 spectra of the sample and 400 spectra of the background were randomly selected. These spectra were used to build a training set with size {22,400 spectra × 96 wavelengths}, containing 16,000 representative spectra of the considered materials (2000 spectra for each material type) and 6400 spectra of the background. The average spectra of each material type calculated from the training set are reported in Figure 1A, while the average and standard deviation spectra of each class are shown in Figure S1 of the Supplementary Material.

The images containing samples of two different materials were used as test images to obtain both a quantitative and qualitative evaluation of the classification models. For the creation of the test set, some represen-

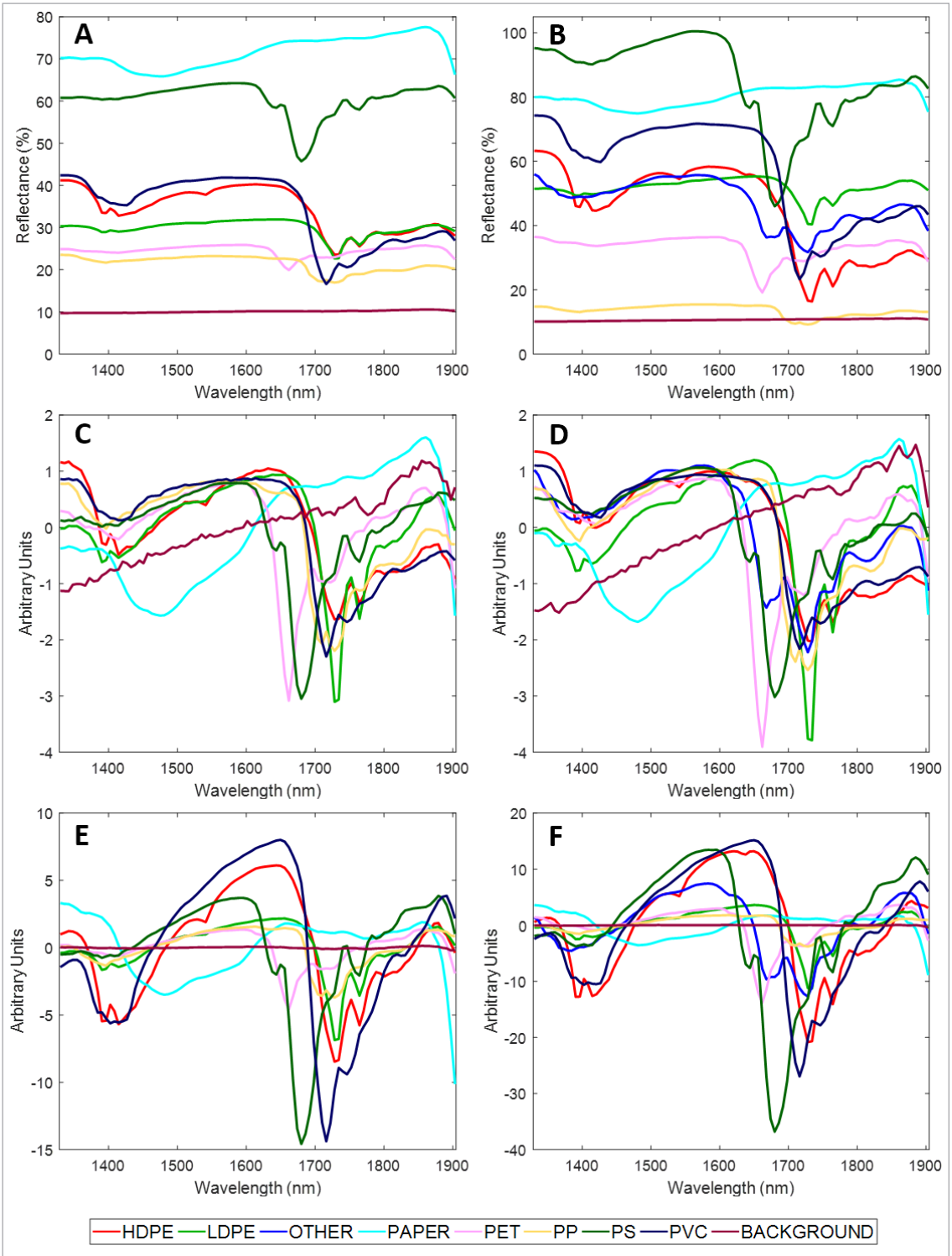

Figure 1. Average spectra of each material type calculated from training and test sets without any pre-processing (A and B), and after SNV (C and D) and detrend (E and F) as data preprocessing.

tative images have been chosen in order to consider all the materials, and Regions of Interest (ROI) were defined on these images. From each ROI, the corresponding spectra were extracted to create a test set with size {21,760 spectra × 96 wavelengths}, containing a library of spectra with known assignment. Figure 1B shows the average spectra of each considered material calculated from the test set, while the corresponding average and

standard deviation spectra are reported in Figure S2 of the Supplementary Material.

Data analysis

The training set was initially analysed by means of PCA considering both standard normal variate (SNV) and detrend as row preprocessing methods. This preliminary analysis allowed similarities and differences between the considered materials to be identified, and both SNV and detrend gave analogous results. The clusters of the different materials observed in the PCA score plots essentially reflected similarities and differences of the chemical structure of the considered polymers. These results were used to develop the structure of the tree-based classification model reported in Figure 2. In particular, PCA highlighted that the class corresponding to the non-recyclable plastic polymers was too heterogenous to be modelled into a single class. For this reason, the spectra belonging to the OTHER class were not used during model calibration, but they were used during the final validation of the classification tree in order to evaluate the ability of Soft PLS-DA algorithm to reject objects belonging to classes not considered in model calibration. The classification tree reported in Figure 2 is composed of five nodes, each corresponding to a classification model calculated with Soft PLS-DA algorithm using the training set.

For each node of the tree, the classification models have been computed considering both *SNV + mean centring* and *detrend + mean centring* as data preprocessing methods.

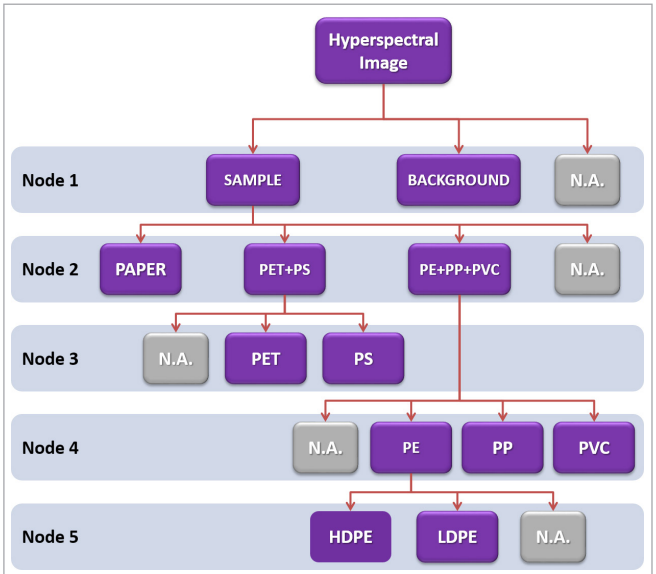

Figure 2. Tree-structured classification models.

The average spectra of each considered material type obtained from both training and test sets preprocessed with SNV and detrend are reported in Figures 1C–F.

Classification performances of each single class were defined using sensitivity (SENS), i.e. the percentage of objects belonging to a given class correctly assigned to the corresponding class, specificity (SPEC), i.e. the percentage of objects correctly rejected from the class model, and efficiency (EFF), i.e. the geometric mean of sensitivity and specificity. Furthermore, for an overall evaluation of the classification quality of each node, the Non-Error Rate (NER) was also considered, which is calculated as the arithmetic mean of the SENS values of the different classes.[33]

The proper number of LVs for the Soft PLS-DA models has been optimised by maximising the NER value in cross-validation. In particular, a customised cross-validation scheme has been adopted, consisting of two deletion groups (i.e., cross-validation groups); during the cross-validation step one model was therefore calculated using the spectra of each deletion group, to predict the class of the spectra of the other deletion group.

Each deletion group was defined in order to include the spectra belonging to all the considered target classes and to keep together all the spectra extracted from the same image.

Furthermore, for each node of the classification tree, the Soft PLS-DA algorithm has also been coupled with a sparse-based variable selection approach in order to identify the relevant variables involved in the classification. In more detail, the outcomes of the sparse PLS-DA (sPLS-DA) algorithm proposed by Lê Cao *et al.*[34–36] were subjected to the same constraints for class assignment described above for Soft PLS-DA, thus resulting in a sparse version of Soft PLS-DA (sparse Soft PLS-DA).

Basically, the main idea of sparse-based methods is to perform variable selection by forcing the model coefficients not bringing useful information to be equal to zero. Sparse methods represent an extension of the corresponding traditional classification or regression methods, where sparsity is achieved by adding a penalty term to the computation of the model coefficients.[37,38] In addition to the number of model components (i.e., LVs), sparse methods also require the level of sparsity to be optimised, which is related to the number of variables whose coefficients are set equal to zero in the model.[36,39]

The sparse-based Soft PLS-DA models were optimised considering a maximum number of LVs equal to 10 and

a number of variables selected for each LV ranging from 4 to 96 (with a step equal to 4). For each node of the classification tree, the best combination between the number of LVs and the number of selected variables was identified by maximising the NER value estimated in cross-validation.

Both Soft PLS-DA and sparse Soft PLS-DA models were validated using the test set described above.

Image elaboration and data analysis were performed using the PLS_Toolbox software (ver. 8.5, Eigenvector Research Inc., USA) and *ad hoc* routines developed in the MATLAB environment (ver. 9.0, The MathWorks, USA). The MATLAB routine to run Soft PLS-DA algorithm is freely downloadable from http://www.chimslab.unimore.it/downloads/.

Results

Classification by Soft PLS-DA

Table 1 and Table 2 report the results obtained by applying Soft PLS-DA algorithm for each node of the classification tree and considering both SNV and detrend as row preprocessing methods. For each model, the classification performances have been evaluated considering, for each single class, the number of not-assigned pixel spectra and SENS, SPEC and EFF values, while the NER values have been calculated as a global measure over all the classes. To maintain a concise presentation, Table 1 shows only the NER values obtained for each node of the classification tree, while Table 2 reports the SENS values of all the considered classes in each node of the tree.

Table 1. NER values obtained with Soft PLS-DA in calibration (CAL), cross-validation (CV) and prediction of the test set (TS) for each node of the classification tree.

	SNV + mean centring				Detrend + mean centring			
	LV	CAL	CV	TS	LV	CAL	CV	TS
NODE 1	3	91.8	87.7	98.4	5	86.1	86.0	84.9
NODE 2	2	94.3	93.6	100.0	5	93.9	92.4	91.1
NODE 3	5	94.7	89.5	100.0	2	97.0	94.7	59.6
NODE 4	4	90.3	90.2	98.9	6	87.7	83.9	98.1
NODE 5	5	90.6	82.9	100.0	2	90.7	85.1	99.9

Table 2. SENS values obtained with Soft PLS-DA in calibration (CAL), cross-validation (CV) and prediction of the test set (TS) for each node of the classification tree.

		SNV + mean centring			Detrend + mean centring		
		CAL	CV	TS	CAL	CV	TS
NODE 1	BACKGROUND	91.6	89.7	98.1	96.9	96.6	100.0
	SAMPLE	91.9	85.7	98.6	75.3	75.5	69.9
NODE 2	PAPER	98.7	98.8	100.0	100.0	99.9	100.0
	PE+PVC+PP	93.5	92.9	100.0	92.3	89.5	96.6
	PS+PET	90.6	89.2	100.0	80.8	76.7	50.2
NODE 3	PET	95.2	85.9	100.0	95.4	97.4	76.2
	PS	94.3	93.2	100.0	98.6	92.1	43.0
NODE 4	PE	89.8	87.7	99.7	86.2	82.8	97.7
	PP	89.5	91.2	97.1	91.6	86.3	99.5
	PVC	91.8	91.8	100.0	85.5	82.6	97.1
NODE 5	HDPE	86.6	84.5	100.0	81.8	76.5	99.8
	LDPE	94.6	81.3	100.0	99.6	93.7	100.0

With few exceptions, SNV preprocessing method allowed better classification performances to be obtained than detrend, both in cross-validation and in prediction of the test set. Indeed, in Node 1, Node 2 and Node 4 the NER values obtained in cross-validation from the models calculated with SNV are higher than those obtained with detrend, and these results were also confirmed from the prediction of the external test set. Concerning Node 3, SNV and detrend led to similar classification performances in cross-validation, but detrend gave a lower NER value in prediction of the test set; in particular a SENS value equal to 43.0 % was obtained for class PS. Therefore, also for Node 3 SNV was the best preprocessing method, since it resulted in a more robust classification model. Conversely, in Node 5 the two preprocessing methods showed similar performances both in cross-validation and in prediction of the test set.

Based on these results, SNV can be identified as the optimal preprocessing method for all the five nodes of the classification tree. Furthermore, it has to be highlighted that the use of the same preprocessing method for each node of the classification tree represents a great advantage from the computational point of view. Indeed, the spectra of the hyperspectral images can be row-preprocessed only once after image acquisition, and the preprocessed spectra can then be used for all the nodes of the classification tree, resulting in a lower computational effort.

Figure S3 of the Supplementary Material shows the regression vectors of the classification models calculated for each node of the tree using SNV as spectral row preprocessing. By comparing Figure S3 with Figure 1, which reports the average spectra of the different material types, it is possible to observe that for each node of the tree the relevant variables, i.e. the variables with highest absolute values of the regression coefficients, generally correspond to the characteristic wavelengths of the materials considered in the specific classification problem.

Classification by Soft PLD-DA and sparse-based variable selection

Since SNV was the optimal row preprocessing method for the whole classification tree, the subsequent variable selection step by means of sparse Soft PLS-DA was performed considering only *SNV + mean centring* for each node.

Table 3 reports the classification results obtained for the five nodes of the classification tree expressed in terms of NER vales as a global measurement of the classification performances of each node, while Table 4 shows the SENS values of each modelled class.

Comparing Table 1 with Table 3 and Table 2 with Table 4, it is possible to observe that, generally, variable selection improved the classification performances and, at the same time, considerably reduced the number of retained variables.

The higher improvement in classification performances was reached in Node 3 with a NER value in cross-validation equal to 94.3 % with respect of a NER value equal to 89.5 % obtained considering the full wavelength

Table 3. NER values obtained with sparse-based variable selection coupled with Soft PLS-DA in calibration (CAL), cross-validation (CV) and prediction of the test set (TS) for each node of the classification tree.

	LV	No. variables	CAL	CV	TS
NODE 1	4	62	92.7	89.1	99.2
NODE 2	2	8	95.6	95.2	100.0
NODE 3	1	20	97.8	94.3	100.0
NODE 4	4	40	90.3	90.5	98.1
NODE 5	2	6	89.6	81.9	100.0

Table 4. SENS values obtained with sparse-based variable selection coupled with Soft PLS-DA in calibration (CAL), cross-validation (CV) and prediction of the test set (TS) for each node of the classification tree.

		CAL	CV	TS
NODE 1	BACKGROUND	92.5	90.5	98.9
	SAMPLE	92.8	87.7	99.6
NODE 2	PAPER	100.0	99.8	100.0
	PE+PVC+PP	94.6	94.6	100.0
	PS+PET	92.3	91.3	100.0
NODE 3	PET	97.4	92.5	100.0
	PS	98.2	96.1	100.0
NODE 4	PE	88.4	87.2	94.7
	PP	91.9	93.1	99.5
	PVC	90.7	91.3	100.0
NODE 5	HDPE	92.9	82.6	100.0
	LDPE	86.3	81.3	100.0

range. In addition, by considering only the 20 selected variables in Node 3, all the test set spectra were correctly assigned.

Node 2 is the node with the lower number of selected variables, retaining only 8 variables out of the 96 original wavelengths. Such a small number of variables enabled an increase in the classification performances for all the three classes modelled in Node 2 (PAPER, PE+PVC+PP and PS+PET), maintaining at the same time the 100.0 % of correct assignments for the test set.

Conversely, in Node 5 variable selection led to a slight decrease of the classification performances in cross-validation; in particular, the SENS value for class HDPE is lower than what obtained with the full wavelength range (82.6 % vs 84.5 %). Actually, it should be considered that Node 5 deals with the classification of HDPE and LDPE, which are derived from the same monomer, and the difference between these two polymers is only related to the degree of branching. Therefore, it is reasonable to assume that, for the discrimination between HDPE and LDPE, the full wavelength range provides more complete information.

Final implementation of the classification tree

Based on the results obtained from the optimisation of each single node, the different classification models have been assembled together in order to obtain the final implementation of the classification tree, which is schematised in Figure 3.

The proposed tree-structured classification model was applied to the test set in order to obtain a quantitative assessment of its classification performances. During the final validation of the model, the spectra belonging to the other types of non-recyclable plastics (class OTHER) were also included in the test set to evaluate the ability of Soft PLS-DA, and of its sparse-based extension, to correctly reject spectra not belonging to the modelled classes.

The results are summarised in Figure 4, which can be seen as a kind of graphical representation of the confusion matrix obtained by applying the classification tree to the test set. Indeed, each point in the graph represents a spectrum of the test set and it is coloured according to the corresponding actual class, while the position of the points on the y-axis is based on the class predicted from the tree-structured classification model.

The same results are also reported in Table 5 in terms of SENS and SPEC values for each considered class, and of NER of the overall tree-structured model.

Generally, satisfactory classification results have been obtained for all the modelled classes, resulting in a NER value equal to 98.4 %. The SENS values are always greater than 90 %, and for PS and PVC all the test set spectra have been correctly predicted.

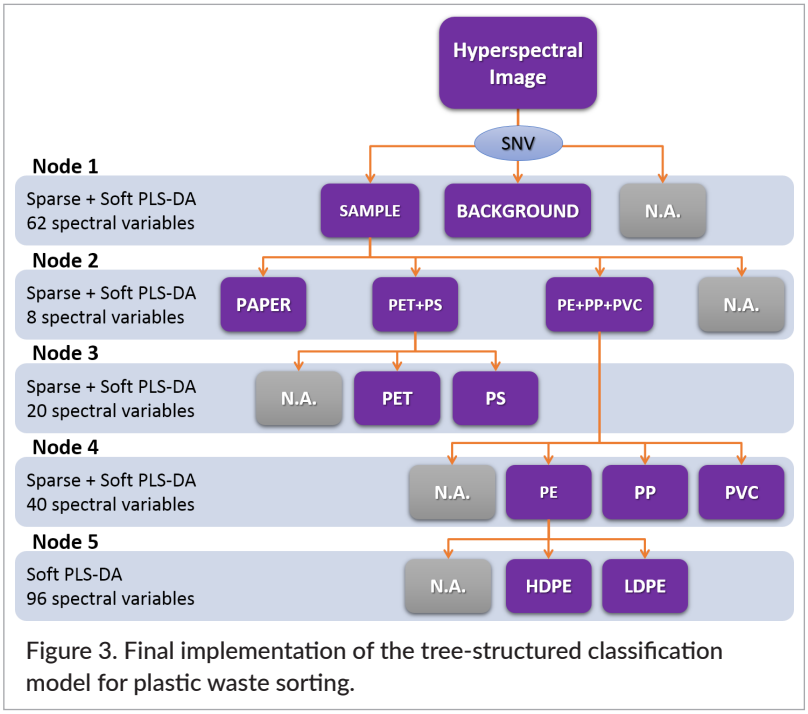

Figure 3. Final implementation of the tree-structured classification model for plastic waste sorting.

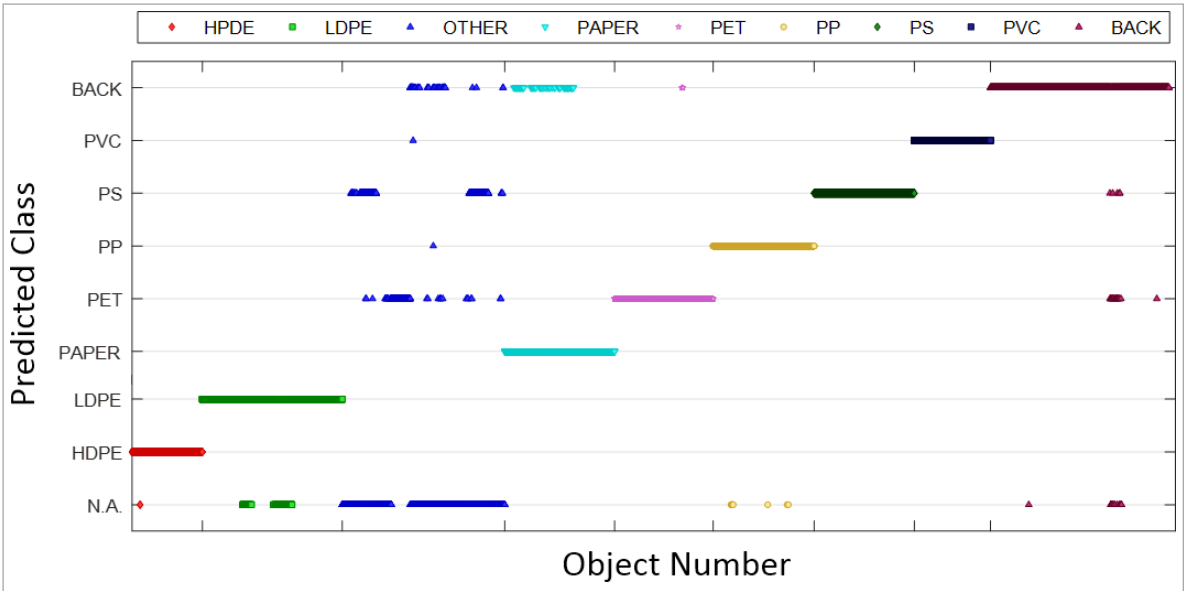

Figure 4. Prediction results obtained from the final implementation of the tree-structured classification model.

Concerning the results expressed in terms of specificity, for all the classes, SPEC values close to 1 were obtained, indicating that the classification tree correctly rejected the great part of spectra not belonging to the considered class. Indeed, Figure 4 shows that the majority of the test set spectra belonging to the category of other plastics (OTHER), which was not included in the classification tree, was correctly rejected from all the considered classes and thus labelled as "not-assigned" (NA). In more detail, 70.9 % of spectra belonging to class OTHER was predicted as NA, while only a minority of these spectra from other plastics was erroneously assigned to PET and PS classes (12.4 % and 14.7 %, respectively).

These results confirm that the proposed tree-structured classification model is able to effectively recognise the spectra of the analysed materials, minimising at the same time possible false positives due to spectra belonging to materials that were not considered during model calibration.

Furthermore, the classification tree was applied to the test images, i.e. to the images containing objects of two different materials, in order to evaluate the classification performances at the pixel level. As an example, Figure 5 shows the prediction images obtained from the hyperspectral images containing PP + PVC (Figure 5A), PAPER + HDPE (Figure 5B), LDPE + OTHER (Figure 5C) and PET + PS (Figure 5D). In order to facilitate the interpretation of the results, the RGB images of the corresponding waste samples have also been included together with the prediction images. In the prediction images, all the pixels predicted as belonging to a defined class have been coloured according to the legend. In particular, grey is associated with those pixels that have not been assigned to any class.

Based on the prediction images, it is possible to observe that the majority of the pixels of each single object are correctly assigned to the corresponding material type. Misclassifications mainly occur at the edges of the objects or in some areas of the background, due to

Table 5. Results in prediction obtained by applying the classification tree to the test set. BACK is the abbreviation for background.

	HDPE	LDPE	PAPER	PET	PP	PS	PVC	BACK
SENS %	99.9	92.1	96.7	99.8	99.5	100.0	100.0	98.9
SPEC %	100.0	100.0	100.0	97.7	100.0	97.4	100.0	99.2
NER %	98.4							

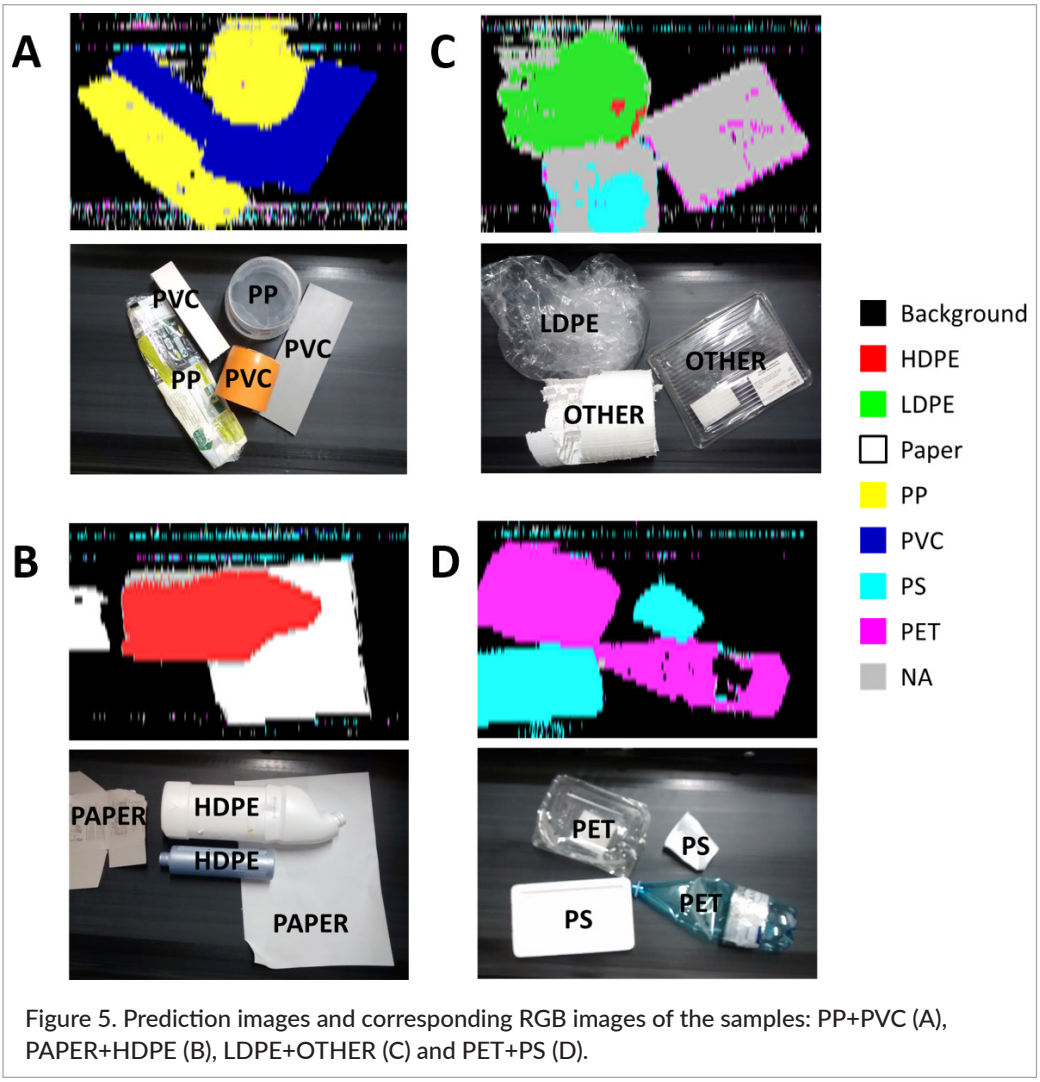

Figure 5. Prediction images and corresponding RGB images of the samples: PP+PVC (A), PAPER+HDPE (B), LDPE+OTHER (C) and PET+PS (D).

the noise caused by specular reflections of the conveyor belt.

In more detail, in Figure 5A and Figure 5D all the pixels of the objects depicted in the images are correctly classified, while in Figure 5B the pixels of the larger HDPE bottle located at its upper edge are not assigned. Considering Figure 5C, the majority of the pixels of the LDPE object are correctly assigned, while some of them are classified as HDPE or not assigned. In the same image, the objects made of plastic polymers not included in the classification model have been globally not assigned to any class.

Conclusions

In practical applications of hyperspectral imaging to classification aims, e.g. in sorting plants, it is necessary to implement classification methods able to effectively handle a large number of classes, to correctly classify samples belonging to the considered categories, and to correctly recognise and reject possible foreign objects.

The present study was aimed at the development of a classification rule for the discrimination of multiple categories through a tree-structured model, in which the classification at each node was performed by Soft PLS-DA, an extension of the PLS-DA algorithm. The basic engine of Soft PLS-DA is the same as PLS-DA, but class assignment is subjected to some additional criteria involving the calculation of further thresholds based on Q residuals and on y predictions. These additional thresholds allow the rejection of unknown samples which are not compliant with the classes of interest.

The proposed approach was tested on a case study related to the discrimination of the different recyclable

plastic polymers that are commonly used for packaging. On the one hand, the use of a tree-structured classification model allowed eight different classes to be efficiently handled, a situation in which a single discrimination step would rarely give satisfactory results. On the other hand, Soft PLS-DA proved to be a flexible algorithm thanks to the possibility of rejecting foreign objects, whose presence is a plausible situation in recycling plants, since it is not possible to completely control the incoming materials. Furthermore, coupling Soft-PLS-DA with a sparse-based variable selection allowed us to improve the classification performances and to decrease the number of spectral variables, reducing at the same time the computational efforts.

The external validation of the classification tree demonstrated the effectiveness of the proposed approach, reaching a NER value equal to 98.4 %. These satisfactory results were also confirmed by the pixel-level prediction performed on a set of test images.

A further improvement of this application will consist in the extension of the prediction from a pixel-level to an object-level approach, by assigning each plastic sample to a defined class based on the class attribution of the majority of its pixels. In the specific case under investigation, this task can be accomplished by coupling the hyperspectral camera with an RGB camera for object shape detection. Indeed, the much higher spatial resolution of RGB imaging could allow to better define the edges of each imaged object, to identify labels partly covering the samples, and to further classify the samples of a given plastic material based on its colour.

Acknowledgements

The authors acknowledge financial support for this research by CNIGROUP – Vision systems business unit through the POR-FESR 2014-2020 project RICIPLA. The authors also wish to thank Philipp Haumann and Edoardo Covertino (CNIGROUP – Vision systems business unit, Alfonsine, Italy) for the help during image acquisition and for technical support.

References

1. A. Gowen, C. O'Donnell, P.J. Cullen, G. Downey and J.M. Frias, "Hyperspectral imaging–an emerging process analytical tool for food quality and safety control", *Trends Food Sci. Technol.* **18(12),** 590 (2007). doi: https://doi.org/10.1016/j.tifs.2007.06.001
2. Y.Y. Pu, Y.Z. Feng and D.W. Sun, "Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: a review", *Compr. Rev. Food Sci F.* **14(2),** 176 (2015). doi: https://doi.org/10.1111/1541-4337.12123
3. C. Ferrari, G. Foca, R. Calvini and A. Ulrici, "Fast exploration and classification of large hyperspectral image datasets for early bruise detection on apples", *Chemometr. Intell. Lab. Syst.* **146,** 108 (2015). doi: https://doi.org/10.1016/j.chemolab.2015.05.016
4. G. ElMasry, F.D. Barbin, D.W. Sun and P. Allen, "Meat quality evaluation by hyperspectral imaging technique: an overview", *Crit. Rev. Food Sci. Nutr.* **52(8),** 689 (2012). doi: https://doi.org/10.1080/10408398.2010.507908
5. K. Sendin, P.J. Williams and M. Manley, "Near infrared hyperspectral imaging in quality and safety evaluation of cereals", *Crit. Rev. Food Sci. Nutr.* **58(4),** 575 (2018). doi: https://doi.org/10.1080/10408398.2016.1205548
6. E. Dufour, "Recent advances in the analysis of dairy product quality using methods based on the interactions of light with matter", *Int. J. Dairy Technol.* **64(2),** 153 (2011). doi: https://doi.org/10.1111/j.1471-0307.2010.00665.x
7. Y. Roggo, A. Edmond, P. Chalus and M. Ulmschneider, "Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms", *Anal. Chim. Acta* **535,** 79 (2005). doi: https://doi.org/10.1016/j.aca.2004.12.037
8. A.A. Gowen, C.P. O'Donnell, P.J. Cullen and S.E.J. Bell, "Recent applications of chemical imaging to pharmaceutical process monitoring and quality control", *Eur. J. Pharm. Biopharm.* **69(1),** 10 (2008). doi: https://doi.org/10.1016/j.ejpb.2007.10.013
9. J.M. Amigo, "Practical issues of hyperspectral imaging analysis of solid dosage forms", *Anal. Bioanal. Chem.* **398,** 93 (2010). doi: https://doi.org/10.1007/s00216-010-3828-z
10. A. Ulrici, S. Serranti, C. Ferrari, D. Cesare, G. Foca and G. Bonifazi, "Efficient chemometric strategies for PET–PLA discrimination in recycling plants using hyperspectral imaging", *Chemometr. Intell. Lab. Syst.* **122,** 31 (2013). doi: https://doi.org/10.1016/j.chemolab.2013.01.001

11. P. Tatzer, M. Wolf and T. Panner, "Industrial application for inline material sorting using hyperspectral imaging in the NIR range", *Real-Time Imaging* **11(2),** 99 (2005). doi: https://doi.org/10.1016/j.rti.2005.04.003

12. R. Palmieri, G. Bonifazi and S. Serranti, "Recycling-oriented characterization of plastic frames and printed circuit boards from mobile phones by electronic and chemical imaging", *Waste Manag.* **34(11),** 2120 (2014). doi: https://doi.org/10.1016/j.wasman.2014.06.003

13. J.M. Amigo, H. Babamoradi and S. Elcoroaristizabal, "Hyperspectral image analysis. A tutorial", *Anal. Chim. Acta* **896,** 34 (2015). doi: https://doi.org/10.1016/j.aca.2015.09.030

14. P. Paclik, R. Leitner and R.P.W. Duin, "A study on design of object sorting algorithms in the industrial application using hyperspectral imaging", *J. Real-Time Image Proc.* **1(2),** 101 (2006). doi: https://doi.org/10.1007/s11554-006-0018-5

15. L. Eriksson, J. Trygg and S. Wold, "PLS-trees®, a top-down clustering approach", *J. Chemometr.* **23(11),** 569 (2009). doi: https://doi.org/10.1002/cem.1254

16. K.H. Liland, A. Kohler and V. Shapaval, "Hot PLS—a framework for hierarchically ordered taxonomic classification by partial least squares", *Chemometr. Intell. Lab. Syst.* **138,** 41 (2014). doi: https://doi.org/10.1016/j.chemolab.2014.07.010

17. W. Song, H. Wang, P. Maguire and O. Nibouche, "Nearest clusters based partial least squares discriminant analysis for the classification of spectral data", *Anal. Chimica Acta* **1009,** 27 (2018). doi: https://doi.org/10.1016/j.aca.2018.01.023

18. S. Chevallier, D. Bertrand, A. Kohler and P. Courcoux, "Application of PLS-DA in multivariate image analysis", *J. Chemometr.* **20(5),** 221 (2006). doi: https://doi.org/10.1002/cem.994

19. M. Barker and W. Rayens, "Partial least squares for discrimination", *J. Chemometr.* **17(3),** 166 (2003). doi: https://doi.org/10.1002/cem.785

20. N.F. Pérez, J. Ferré and R. Boqué, "Calculation of the reliability of classification in discriminant partial least-squares binary classification", *Chemometr. Intell. Lab. Syst.* **95(2),** 122 (2009). doi: https://doi.org/10.1016/j.chemolab.2008.09.005

21. P. Oliveri, "Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues–A tutorial", *Anal. Chim. Acta* **982,** 9 (2017). doi: https://doi.org/10.1016/j.aca.2017.05.013

22. S. Wold, "Pattern recognition by means of disjoint principal components models", *Pattern Recognit.* **8(3),** 127 (1976). doi: https://doi.org/10.1016/0031-3203(76)90014-5

23. Y. Zheng, J. Bai, J. Xu, X. Li and Y. Zhang, "A discrimination model in waste plastics sorting using NIR hyperspectral imaging system", *Waste Manag.* **72,** 87 (2018). doi: https://doi.org/10.1016/j.wasman.2017.10.015

24. G. Bonifazi, G. Capobianco and S. Serranti, "A hierarchical classification approach for recognition of low-density (LDPE) and high-density polyethylene (HDPE) in mixed plastic waste based on short-wave infrared (SWIR) hyperspectral imaging", *Spectrochim. Acta A* **198,** 115 (2018). doi: https://doi.org/10.1016/j.saa.2018.03.006

25. P. Geladi and B.R. Kowalski, "Partial least-squares regression: a tutorial", *Anal. Chim. Acta* **185,** 1 (1986). doi: https://doi.org/10.1016/0003-2670(86)80028-9

26. A. Höskuldsson, "PLS regression methods", *J. Chemometr.* **2(3),** 211 (1988). doi: https://doi.org/10.1002/cem.1180020306

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

28. O.Y. Rodionova, A.V. Titova and A.L. Pomerantsev, "Discriminant analysis is an inappropriate method of authentication", *TrAC Trends Anal. Chem.* **78,** 17 (2016). doi: https://doi.org/10.1016/j.trac.2016.01.010

29. A.L. Pomerantsev and O.Y. Rodionova, "Multiclass partial least squares discriminant analysis: Taking the right way—A critical tutorial", *J. Chemometr.* **32,** 3030 (2018). doi: https://doi.org/10.1002/cem.3030

30. C. Botella, J. Ferré and R. Boqué, "Classification from microarray data using probabilistic discriminant partial least squares with reject option", *Talanta* **80,** 321 (2009). doi: https://doi.org/10.1016/j.talanta.2009.06.072

31. J.E. Jackson and G.S. Mudholkar, "Control procedures for residuals associated with principal component analysis", *Technometrics* **21,** 341 (1979). doi: https://doi.org/10.1080/00401706.1979.10489779

32. ASTM D7611 / D7611M-13e1, *Standard Practice for Coding Plastic Manufactured Articles for Resin Identification*. ASTM International, West Conshohocken, PA, USA (2013). https://www.astm.org

33. D. Ballabio, F. Grisoni and R. Todeschini, "Multivariate comparison of classification performance measures", *Chemometr. Intell. Lab. Syst.* **174,** 33 (2018). doi: https://doi.org/10.1016/j.chemolab.2017.12.004

34. K.A. Lê Cao, S. Boitard and P. Besse, "Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems", *BMC bioinformatics* **12,** 253 (2011). doi: https://doi.org/10.1186/1471-2105-12-253

35. E. Szymańska, E. Brodrick, M. Williams, A.N. Davies, H.J. van Manen and L.M. Buydens, "Data size reduction strategy for the classification of breath and air samples using multicapillary column-ion mobility spectrometry", *Anal. Chem.* **87(2),** 869 (2015). doi: https://doi.org/10.1021/ac503857y

36. R. Calvini, A. Ulrici and J.M. Amigo, "Practical comparison of sparse methods for classification of Arabica and Robusta coffee species using near infrared hyperspectral imaging", *Chemometr. Intell. Lab. Syst.* **146,** 503 (2015). doi: https://doi.org/10.1016/j.chemolab.2015.07.010

37. P. Filzmoser, M. Gschwandtner and V. Todorov, "Review of sparse methods in regression and classification with application to chemometrics", *J. Chemometr.* **26,** 42 (2012). doi: https://doi.org/10.1002/cem.1418

38. M.A. Rasmussen and R. Bro, "A tutorial on the Lasso approach to sparse modeling", *Chemometr. Intell. Lab. Syst.* **119,** 21 (2012). doi: https://doi.org/10.1016/j.chemolab.2012.10.003

39. R. Calvini, J.M. Amigo and A. Ulrici, "Transferring results from NIR-hyperspectral to NIR-multispectral imaging systems: a filter-based simulation applied to the classification of Arabica and Robusta green coffee", *Anal. Chimica Acta* **967,** 33 (2017). doi: https://doi.org/10.1016/j.aca.2017.03.011

Highlighting Young Investigators

In this series we are highlighting the contributions of young researchers who are leading exciting new research paths in Spectral Imaging. Contributors, primarily PhD students or Postdoctoral researchers under the age of 40, are selected by the editorial board, based on having published or presented substantially novel research in Spectral Imaging. Nominations will be accepted and should include a statement (no longer than one page) on the researcher's novel research in Spectral Imaging, along with a CV. Please send nominations to *JSI*'s Editor-in-Chief, Aoife Gowen (aoife.gowen@ucd.ie).

Dr Rosalba Calvini was born in Genova (Italy) in 1988. She obtained her PhD (cum laude) in Agri-Food Sciences Technologies and Biotechnologies in 2017, defending a thesis titled "Chemometric tools for food characterization through RGB and hyperspectral imaging". Since 2017, she has been a post-doctoral researcher at the Department of Life Sciences, University of Modena and Reggio Emilia.

Her research interests include the application and development of chemometric algorithms for hyperspectral and digital image analysis with a focus on variable selection and data dimensionality reduction, the use of spectroscopy (mainly NIR spectroscopy) for the characterisation of raw materials or complex food matrices, and the application of data fusion strategies for the joint analysis of data deriving from different analytical instrumentations.

She has co-authored multiple papers in international peer-reviewed scientific journals and presented the results of her research activities in several national and international conferences. She has been involved in teaching chemometrics applied to image analysis in PhD and professional courses. She has been co-supervisor of multiple thesis for the MSc in Food Control and Security (University of Modena and Reggio Emilia) and of one PhD thesis in Agri-Food Sciences, Technologies and Biotechnologies.