Chemometrics and Intelligent Laboratory Systems 277 (2026) 105817

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

NIR hyperspectral imaging for the identification of damage caused by *Halyomorpha halys* on pears: development of classification models

Veronica Ferrari [a], Rosalba Calvini [a,b,*], Camilla Menozzi [a], Elena Costi [a], Peter Offermans [c], Lara Maistrello [a], Alessandro Ulrici [a,b]

[a] University of Modena and Reggio Emilia, Department of Life Sciences, Pad. Besta, Via Amendola, 2, Reggio Emilia, 42122, Italy
[b] Consorzio Interuniversitario Nazionale per la Scienza e Tecnologia dei Materiali (INSTM), Via Giusti, 9, Firenze, 50121, Italy
[c] IMEC OnePlanet, Bronland 10, Wageningen, the Netherlands

ARTICLE INFO

Keywords:
Hyperspectral imaging
Fruit puncture detection
Post-harvest sorting
Multivariate classification
Soft PLS-DA
Sparse methods

ABSTRACT

Halyomorpha halys, the Brown Marmorated Stink Bug (BMSB), has become a major threat for pear orchards, causing severe economic losses throughout Southern Europe. Italy, one of the leading European pear producers, has experienced a dramatic reduction in yields of economically relevant cultivars over the past decade due to extreme climate events and outbreaks of pests and pathogens.

To address this issue, Near Infrared Hyperspectral Imaging (NIR-HSI) is proposed as a non-destructive technique to automatically discard damaged fruits. The study was performed on organic pears (*cv*. Abate Fétel and *cv*. Williams) harvested over two consecutive years.

Following an automated annotation procedure that identified punctured and sound areas of the fruits, representative spectra of both *Punctured* and *Sound* classes were extracted to build pixel-level classification models. Classification models were developed using Soft Partial Least Squares Discriminant Analysis (Soft PLS-DA), a hybrid approach that combines the advantages of both discriminant and class modelling techniques, and its sparse variant (s-Soft PLS-DA), which incorporates variable selection to achieve a parsimonious solution more suitable for online post-harvest sorting systems.

The prediction images were elaborated considering an additional decision criterion based on the minimum detectable defect area, enabling image-level classification. Besides being advantageous for automated sorting systems, image-level classification allowed a comprehensive assessment of models' classification performances towards modelled and unmodelled (i.e., damage not ascribable to BMSB) classes.

1. Introduction

In recent decades, increasing anthropogenic activities eased the spread of invasive insect pests capable of altering agroecosystems and compromising agri-food production, leading to substantial economic losses [1,2]. Among these, *Halyomorpha halys*, commonly known as the Brown Marmorated Stink Bug (BMSB), is particularly noteworthy. Native to East Asia, it rapidly spread to Europe and North America, becoming a significant agricultural threat. This highly polyphagous pest damages a wide range of crops, including fruits, vegetables, and ornamental plants. Severe outbreaks have been associated with yield losses from 60% up to 100%, causing millions of dollars in economic damage to fruit and horticultural production in Europe and North America [3–5].

BMSB feeds using its piercing-sucking mouthparts, causing deformities, discoloration, surface lesions, and internal tissue damage in fruits [6–8]. In Southern Europe, favourable climatic conditions and high crop density have supported large pest populations [9]. Northern Italy has been particularly affected by BMSB infestations, especially its pear orchards, where this pest represents a relevant agricultural issue. In 2019 alone, economic losses attributed to BMSB were estimated at €590 million for fruit orchards in Northern Italy, including about €356 million losses in the pear sector, with yield reductions exceeding 80% in the most heavily infested orchards [10].

Italy has historically been the leading European country for pear production, with an average annual yield of approximately 707,000

This article is part of a special issue entitled: CCM2025 published in Chemometrics and Intelligent Laboratory Systems.

* Corresponding author. University of Modena and Reggio Emilia, Department of Life Sciences, Pad. Besta, Via Amendola, 2, 42122, Reggio Emilia, Italy.
E-mail address: rosalba.calvini@unimore.it (R. Calvini).

https://doi.org/10.1016/j.chemolab.2026.105817
Received 17 March 2026; Received in revised form 11 June 2026; Accepted 16 July 2026
Available online 20 July 2026
0169-7439/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

tons, ranking third worldwide. However, an increased activity of pathogens and pests such as BMSB together with frequent frosts at fruit set and drought during summer have reduced the production to approximately 343,000 tons over the past three years, with only 22,000 ha of cultivated land remaining, and constantly decreasing [11,12]. In Northern Italy, the production is predominantly concentrated in Lombardy, Emilia-Romagna, and Veneto, where several cultivars of economic relevance are grown, some regulated by the European Protected Geographical Indication (PGI) designation [13]. In particular, Abate Fétel and Williams are among the most widely cultivated cultivars in the Emilia-Romagna region, accounting for approximately 37% and 27% of national production in 2019, respectively [14]. The importance of these cultivars is also characterized by their strong link with the territory, guaranteed by the "*Pere dell'Emilia-Romagna*" PGI designation [15].

Damage caused by BMSB during the early stages of fruit development is easily visible, since it often determines fruit deformation. Conversely, BMSB punctures occurring in the late stages typically result in internal suberification and necrosis of the pulp, producing brown, corky tissue beneath an externally undamaged skin, and are therefore not detectable through visual inspection [6]. Moreover, the feeding activity of BMSB has been associated with increased susceptibility to pathogens including yeast, *Monilinia spp*. and other fungi, which may lead to fruit rot [5,7,16,17]. As a result, internal defects caused by BMSB late damage severely reduce fruit quality and commercial value, while remaining difficult to detect through visual inspection before fruit consumption or deterioration during post-harvest [5].

Currently, post-harvest sorting practices mainly rely on visual inspection or computer vision systems, which are unable to detect the internal injuries caused by BMSB late feeding as they are not visible at the naked eye. At the same time, Near Infrared (NIR) spectroscopy is widely used to assess fruit quality and safety thanks to the possibility of characterising fruits chemical properties [18,19]. However, traditional spectroscopic techniques do not provide information related to the spatial distribution of chemical compounds, therefore they are not useful in the detection of localized defects such as punctures [20].

In this context, NIR spectral imaging represents a viable alternative to conventional computer vision systems operating in the visible range, enabling rapid, non-destructive, and efficient characterization of the chemical and physical properties of food products. Hyperspectral imaging (HSI) combines the strengths of spectroscopy and imaging, allowing the visualization of spatial distribution of the chemical features across a sample's surface [21,22]. This makes NIR-HSI especially suitable for detecting defects in heterogeneous food matrices, as it allows both the assessment of chemical composition and its spatial variation on the sample surface [20,23–31]. However, HSI data-richness is a double-edged sword: on the one hand, it allows a detailed sample characterization while on the other hand, it poses challenges in data handling, storage, and analysis [32,33]. In this context, Multivariate Image Analysis plays a crucial role in addressing the curse of dimensionality and extracting relevant information for the development of robust classification or calibration models able to predict qualitative and quantitative properties of food products.

Several studies have demonstrated the potential of NIR-HSI for detecting bruises, fungal infections and quality defects in fruits [25,34–36]. However, the detection of internal injuries caused by piercing-sucking insects such as BMSB has received limited attention, despite their increasing economic impact. Within this frame, we evaluated NIR-HSI as a possible post-harvest sorting system method for the early detection of BMSB under-peel damage on pears, which is not detectable using cameras working in the visible range. To our knowledge, no previous studies have investigated NIR-HSI to identify internal damage caused by insect punctures on pears.

For the practical application of NIR-HSI systems, it is necessary to develop supervised classification models able to process the hyperspectral images and provide a class assignment for each pixel (i.e., pixel-level classification). To this aim, the identification in the hyperspectral images of Regions of Interests (ROIs) of both sound and damaged areas is a crucial aspect affecting the effectiveness and robustness of the classification models [37,38]. This procedure is often referred to as image annotation.

In a previous study [39], we developed an innovative approach for the automated and objective annotation of BMSB punctured areas. The proposed method involved image dimensionality reduction and image-level classification coupled with spatial features selection.

In the present study, starting from annotated ROIs of punctured and sound areas of the fruits, representative spectra belonging to both classes were extracted to build pixel-level classification models. To achieve a parsimonious solution suitable for the implementation of online post-harvest sorting systems, variable selection was applied to retrieve only the most relevant spectral regions able to distinguish sound areas of the fruits from those affected by BMSB punctures. Indeed, in future developments, the selected spectral regions can be implemented in multispectral cameras, which are more suitable to be used in practical implementations of post-harvest sorting systems as they are faster, cheaper and their optical components are highly robust [40,41].

Based on these considerations, the main objectives of our study were: i) to model the spectral differences between sound and punctured areas; ii) to obtain an interpretable model leading to the selection of the most important spectral variables; and iii) to have the possibility of handling damage types different from those ascribable to BMSB. These requirements led us to the selection of Soft Partial Least Squares Discriminant Analysis (Soft-PLS-DA) algorithm and its sparse variant, sparse Soft PLS-DA (s-Soft PLS-DA), for the development of the classification models [42]. Soft PLS-DA is a soft discriminant algorithm that enables the rejection of spectra that do not belong to any of the modelled classes, thus offering a more robust and flexible solution in sorting systems. Moreover, s-Soft PLS-DA incorporates sparse-based variable selection to model development, allowing to select relevant spectral regions able to discriminate between punctured and sound areas.

Finally, Soft PLS-DA and s-Soft PLS-DA [42] classification models were applied to all the acquired images and a decision criterion was defined to provide classification at the image-level. The criterion was applied to all the prediction images, enabling a global evaluation of models' classification performances.

2. Materials and methods

2.1. Field experimental protocol

Pear samples belonging to Abate Fétel and Williams cultivars were harvested in an organic orchard located in Carpi (Modena, Italy) during summer 2022 and summer 2023.

In both years, 40 tree branches belonging to different plants located in diverse areas of the orchard were considered. At fruit set, the branches were covered with cylindric inclusion cages made to protect the fruits from uncontrolled biotic and abiotic adversities [6,43].

The same day of commercial harvest, 20 inclusion cages were selected and 3 BMSB specimens were manually placed inside each one of the selected cages. In this manner, about half of the fruits were exposed to BMSB feeding for about 3 – 5 days, while the remainder fruits were considered as control samples. Following this procedure, a total of 156 and 160 fruits were collected for *cv.* Abate Fétel and *cv.* Williams, respectively.

The reader is refereed to Ref. [39] for further insights about the experimental protocol followed in the field.

2.2. Image acquisition and elaboration

Image acquisition was performed at 8 subsequent times to monitor the evolution of fruit damage over time, from harvest day (T1) to five weeks later (T2-T8), following the schedule reported in Table S1. Between the different acquisition sessions, the fruits were stored at

refrigerated temperatures of 0 – 2°C to simulate post-harvest storage conditions [44,45].

The hyperspectral camera used for the acquisitions was a line-scanning system (SnapScan SWIR, IMEC One Planet, The Netherlands) working in the 1156 – 1674 nm spectral range with 5 nm spectral resolution (100 channels) [46], as described in detail in Ref. [39]. The images were acquired considering a lens aperture equal to f/2.8 and an integration time equal to 2 ms. Four small halogen lamps with diffusors were utilized for the homogeneous illumination of the samples. The hyperspectral camera used in this study integrates an internal sensor motion mechanism, allowing static image acquisition without external motion of either the camera or the sample. This set up enabled the capture of a full hyperspectral image in approximately 1 s [46]. The acquired hyperspectral images have spatial dimensions of 492 × 820 pixels, with a spatial resolution of 0.2 mm.

To cover the whole pear surface during acquisition, the investigated fruits were divided into four sections (A – D) and each section was imaged, obtaining four hyperspectral images for each fruit.

As previously mentioned, damage due to BMSB punctures is not visible to the naked eye on the intact fruits' surface as it affects the fruit pulp. Therefore, to verify the actual presence of damage it was necessary to peel the fruits after image acquisition.

For this reason and considering the high number of collected fruits, the pears were randomly divided into two groups: the first group (Group 1) composed of fruits to be imaged at each acquisition time and peeled at T8, and a second group of fruits (Group 2) to be imaged at specific acquisition times and peeled at intermediate time points.

In 2022, for each cultivar, 10 control and 10 exposed fruits were assigned to Group 1, while all remaining fruits were assigned to Group 2. In 2023 the same subdivision was carried out for *cv.* Williams fruits, while for *cv.* Abate Fétel Group 1 included only 6 control and 6 exposed fruits due to the lower number of available samples (Table S2 of Supplementary Material).

At each acquisition time, hyperspectral images were acquired of all fruits belonging to Group 1 and of a randomly selected subset of 5 control and 5 exposed fruits belonging to Group 2. After image acquisition, approximately half of the selected fruits from Group 2 were peeled to inspect the presence of damage, while the remaining fruits were stored and subsequently imaged at later acquisition times. Not all fruits in Group 1 could be monitored up to T8 since they reached an over-ripening stage; consequently, some fruits in Group 1 were only acquired until T6 or T7 based on fruit conditions.

Over the two harvest years, 3644 hyperspectral images were acquired, corresponding to 1680 images of Abate Fétel fruits and 1964 images of Williams fruits. The storage space of the whole dataset was equal to 3.1 TB.

Due to adverse climate conditions and the activity of other biotic agents, neither all the control fruits were sound, nor the fruits exposed to BMSB actually showed damage. Therefore, an expert visual inspection was necessary to divide the acquired images into the following categories:

- images of sound sections of control fruits (S–C), corresponding to the images of fruit sections of control samples without any sign of damage;
- images of damaged sections of control fruits (D-C), corresponding to the images of fruit sections of control samples showing damage;
- images of damaged sections of exposed fruits (D-E), corresponding to the images of fruit sections of samples exposed to BMSB with damage;
- images of sound sections of exposed fruits (S-E), corresponding to the images of fruit sections of samples exposed to BMSB but not showing any kind of damage.

Table S2 of Supplementary Material summarises the number of images belonging to the different categories for both Abate Fétel and Williams cultivars. It has to be highlighted that each control fruit did not necessarily generate 4 S–C sections since one or more sections of the same fruit could exhibit damage from other agents and thus labelled as D-C. Similarly, each exposed fruit did not contribute 4 D-E sections since, usually, damaged areas were not located in all four sections. Only the sections of exposed fruits showing damage were labelled as D-E, while the remaining sections were labelled as S-E.

Based on visual inspection, the fruits showing damage were additionally classified into three categories based on damage type:

- Type 1: mild damage of unknown origin (e.g., superficial signs of the fruit pulp);
- Type 2: damage ascribable to BMSB punctures (e.g., suberifications);
- Type 3: severe damage not caused by BMSB punctures (e.g., moulds or other diseases).

Fruit sections showing multiple damage types were assigned to the most severe damage category observed. For example, sections showing both Type 1 and Type 2 damage were labelled as Type 2, while sections exhibiting Type 3 damage together with either Type 1 or Type 2 damage were labelled as Type 3.

Table S3 of Supplementary Material reports the number of images belonging to D-E class with Type 2 damage, which will be referred to as P-E images (i.e., images of punctured sections of fruits exposed to BMSB).

Preliminary image elaboration steps were necessary to remove the background and mask the fruit area. These steps involved thresholding procedures based on Principal Component Analysis (PCA) and morphological erosion, as further detailed in the previous paper [39].

2.3. Development of pixel level classification models

2.3.1. Annotation of ROIs belonging to punctured areas

The development of pixel-level classification models able to discriminate between pixel spectra of sound and punctured areas started with the construction of a library of reference spectra representative of both classes. This task requires the availability of the ground truth, i.e., the exact location of the punctured regions on P-E fruits, quite challenging to achieve in this scenario.

Indeed, to mimic a real situation, the pears were exposed to BMSB directly in the field, therefore it was possible to verify the presence of the punctures and locate them only after peeling the fruits. In addition, standard thresholding approaches for ROIs selection, like e.g., PCA score values of the hyperspectral images, were not easily applicable due to the slight differences between sound and damaged areas, the irregular shape of the punctures, and the high number of acquired images.

To solve this issue, in a previous study we proposed an innovative approach for image annotation that allows the direct comparison of a high number of images altogether and the automated identification of ROIs related to target spatial features [39]. The detailed description of the image annotation methodology is reported in Ref. [39], while this section is intended to provide a concise summary of the key steps.

The image annotation procedure was carried out by analysing separately the images belonging to the two pear cultivars due to their considerably different physical characteristics (i.e., shape, dimensions, peel russeting). For each cultivar, the annotation models were developed between 2022 and 2023 harvest campaigns. In particular the S–C and P-E images of 2022 samples were divided into training (TR) set and test (TS) set images, as reported in Table 1. The subdivision of the images into TR and TS was done by keeping into the same set the images related to the same fruit sample, also across different acquisition times. The TR set images were used to develop the models for ROIs annotation and the TS set images were used for validation.

The proposed annotation algorithm followed several subsequent steps. Firstly, data dimensionality reduction (DDR) was performed using the Common Space Hyperspectrograms (CSH) approach, which converts

Table 1
S–C and P-E images collected in 2022 and 2023 for cv. Abate Fétel and cv. Williams. For P-E images, "P-E Total" refers to the overall number of P-E images acquired, while "P-E Annotated" refers to the number of P-E images for which the punctured regions were successfully identified by the image annotation algorithm proposed in Ref. [39]. The TR and TS partition of 2022 images reflects the use of TR images for the development of the image annotation algorithm and of TS images for its validation, while 2023 images were used as a further validation of the annotation procedure.

	# images	Abate Fétel			Williams		
		S–C	P-E Annotated	P-E Total	S–C	P-E Annotated	P-E Total
2022	TR	246	82	136	112	22	39
	TS	153	58	85	130	10	15
	Total	**399**	**140**	**221**	**242**	**32**	**54**
2023	**Total**	**250**	**70**	**101**	**340**	**240**	**298**

each image into a signal, named hyperspectrogram, built by merging in sequence the frequency distribution curves of quantities (i.e., score, Hotelling T^2 and Q residuals vectors) obtained from a global PCA model common to the considered images [47]. In particular, each hyperspectral image was initially unfolded into a two-dimensional data matrix, where the rows correspond to the pixels retained after background removal and the columns to the spectral variables. The unfolded hypercubes were preprocessed row-wise using Multiplicative Scatter Correction, MSC (median spectrum as reference) for *cv.* Abate Fétel and linear-detrend for *cv.* Williams, and then centred according to the global average spectrum of the corresponding cultivar. For each cultivar, a common PC space was obtained by applying Singular Value Decomposition (SVD) to the kernel variance-covariance matrix, calculated as the sum of the single variance covariance matrices of all the corresponding TR images, and 7 PCs were retained for both cultivars. Subsequently, the unfolded and preprocessed hypercubes were projected onto the common PC space of the corresponding cultivar, and the score vectors, Q residuals and Hotelling's T^2 values were calculated for each image. Finally, the CSH signal of each image was obtained by merging in sequence the frequency distribution curves of the seven score vectors, of Q residuals and of T^2 values. This procedure resulted in two distinct CSH data matrices, one for Abate Fétel and one for Williams samples.

Then, interval PLS-DA (iPLS-DA) models were calculated using the CSH datasets by considering different interval sizes, with the aim of discriminating images of sound and punctured fruits, and selecting the intervals of hyperspectrograms variables that are mostly related to the presence of punctures. Finally, for both cultivars the variables selected most frequently by the different iPLS-DA models were reconstructed back into the original image domain. Indeed, the selected hyperspectrograms variables correspond to groups of pixels characterised by score values falling within a specific interval and more likely related to P-E class. These pixels were visualised back into the original 2-D refolded image domain and considered as annotations of ROIs of punctured areas. A more detailed description of the image annotation algorithm can be found in Ref. [39].

Concerning 2022 harvest campaign, 140 out of 221 P-E images of *cv.* Abate Fétel samples and 32 out of 54 P-E images of *cv.* Williams samples were correctly annotated considering the image annotation procedure described above.

Then, the same annotation procedure was applied to the P-E images acquired in 2023, leading to 70 out of 101 P-E images of *cv.* Abate Fétel samples and 240 out of 298 P-E images of *cv.* Williams samples being successfully annotated. Further details about the outcomes of image annotation are reported in Table 1. It should be noted that in Ref. [39] the annotation procedure was reported for samples belonging to *cv.* Williams, while the present syudy includes both cultivars. For completeness, Table 1 presents the annotation outcomes obtained for both cultivars.

2.3.2. Spectra selection and dataset structure

After annotating the ROIs of punctured regions as described in Section 2.3.1, representative spectra of both punctured and sound areas were randomly selected from S–C and annotated P-E hyperspectral images of both cultivars. The spectra selection phase is crucial in determining the representativeness of spectral signatures for punctured and sound areas, which holds deeply the robustness and reliability of the classification models.

The images belonging to other categories (i.e., D-C, S-E, D-E, not annotated P-E) were not considered in the development of pixel-level classification models, but they were used in a second moment to evaluate their robustness (*see* Section 2.4).

The images were initially divided into TR and TS images. For the 2022 samples of both cultivars, the same TR/TS partition considered for the annotation procedure was applied. To increase the number and representativeness of the samples, also the 2023 images were divided into TR and TS images, as reported in Table 2. Also in this case, images belonging to the same fruit sample, even when acquired at different acquisition times, were kept in the same set.

The spectra signatures for the *Sound* class were selected exclusively from hyperspectral images of S–C fruit sections, while the representative spectra of the *Punctured* class were extracted only from ROIs of punctured areas in the annotated P-E images, generated using the annotation procedure described in Section 2.3.1. Prior to the extraction of representative spectra, each S–C and P-E annotated hyperspectral image was unfolded into a two-dimensional data matrix, with rows corresponding to individual pixels and 100 columns corresponding to the spectral channels. To enable the random selection of spectra, for S–C images only the spectra belonging to the fruit and retained after background removal were kept, while for P-E images only the spectra belonging to the annotated ROIs of punctured areas were considered.

As mentioned in Section 2.2, fruits in Group 1 were imaged at all the eight acquisitions times, while fruits in Group 2 were acquired only at specific time points. As a result, Group 1 fruits had a much higher number of images compared to Group 2 fruits. Therefore, extracting a fixed number of randomly selected spectra from each S–C and P-E image would have led to an over representation of Group 1 fruits and an under representation of Group 2 fruits. To address this issue, we fixed the total number of spectra to be extracted from each S–C or P-E fruit, and then we calculated accordingly the number of spectra to be randomly selected from each corresponding hyperspectral image.

Furthermore, the number of P-E images was much smaller than the number of S–C images and this difference was considered when constructing the dataset of representative spectra, with the aim of obtaining a balanced representation of the *Sound* and *Punctured* classes in terms of number of extracted spectra.

Based on these considerations, a higher number of spectra were extracted from each P-E fruit compared to S–C fruits. More in detail, for each S–C fruit about 260 spectra were randomly selected from the corresponding hyperspectral images, while for each P-E fruit about 460 spectra were randomly selected from the corresponding ROIs in the images.

When only one image was available per fruit, 260 and 460 spectra were selected for S–C and P-E images, respectively. Conversely, for fruits with multiple images (e.g., those belonging to Group 1), the number of spectra to extract from each image was determined by dividing the total number of spectra per fruit by the number of images of that fruit.

It has to be considered that, in some cases, the size of the ROIs annotated on P-E fruits was smaller than 460 spectra; in such cases, all the spectra belonging to the annotated ROI were selected.

More detailed information about the subdivision of fruits, images and spectra in TR and TS for each cultivar can be found in Table 2.

Fig. 1 reports the average spectrum of the selected pixel spectra related to the *Punctured* and *Sound* classes belonging to the TR for both cultivars.

Table 2
S–C and P-E images, along with the corresponding number fruits and the number of selected spectra considered for the creation of the TR and TS set, subdivided by harvest year and cultivar. TR and TS image subdivision was performed at the fruit-level, and the fruit count was based on control fruits with at least one S–C image and exposed fruits with at least one annotated P-E image. It should be noted that a single control fruit did not necessarily generate four S–C sections and a single exposed fruit did not necessarily generate four P-E sections, as specified in Section 2.2.

			S–C			P-E annotated		
			# images	# fruits	# spectra	# images	# fruits	# spectra
Abate Fètel	TR	2022	246	28	6218	82	20	6274
		2023	166	9	1954	47	5	1576
		Total	**412**	**37**	**8172**	**129**	**25**	**7850**
	TS	2022	153	17	5314	58	14	6000
		2023	84	7	2200	23	3	1158
		Total	**237**	**24**	**7514**	**81**	**17**	**7158**
Williams	TR	2022	112	10	2990	22	5	2542
		2023	228	17	5117	159	9	5571
		Total	**340**	**27**	**8107**	**181**	**14**	**8113**
	TS	2022	130	25	5348	10	4	2263
		2023	112	12	2600	81	9	5114
		Total	**240**	**37**	**7948**	**91**	**13**	**7377**

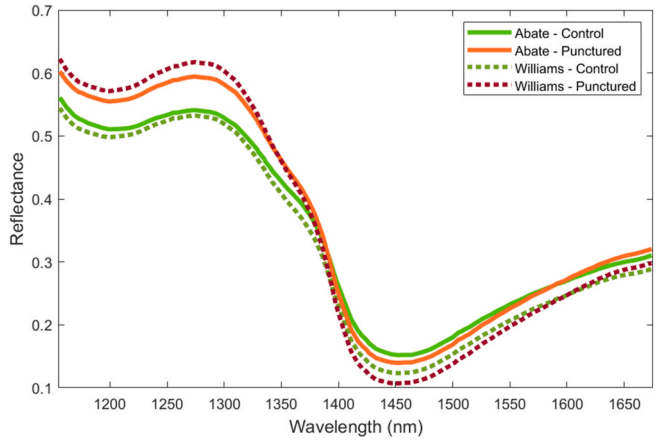


Fig. 1. Average raw spectra representative for the Punctured and Sound classes belonging to the TR sets related to cv Abate Fétel and cv. Williams.

2.3.3. Model calculation and validation

The datasets of representative spectra of *Sound* and *Punctured* classes were used to develop pixel-level classification models using Soft PLS-DA algorithm. Coherently with the annotation procedure described in Section 2.3.1, separate classification models were developed for *cv.* Abate Fètel and *cv.* Williams, as they show considerably different physical characteristics.

The choice of Soft PLS-DA algorithm for classification was motivated by the intrinsic nature of the problem. The previous study related to ROIs annotation of punctured areas evidenced the existence of only subtle spectral differences between sound and damaged tissues. Consequently, the within-class variability of *Sound* and *Punctured* categories is much higher than the variability between the classes themselves, leading to a strong overlap in the feature space. In such scenarios, class modelling techniques often provide limited performances as they focus on modelling class variability rather than maximizing class separation [48]. Nevertheless, class modelling methods offer the important advantage of being able to identify outliers or foreign objects belonging to classes not included in the calibration. Conversely, traditional discriminant approaches such as PLS-DA are specifically designed to maximise the separation between classes, resulting more successful in the presence of overlapping classes. However, these methods are not effective in identifying anomalous objects not belonging to any of the modelled classes.

Soft PLS-DA is a soft discriminant algorithm that combines the advantages of discriminant analysis and class modelling approaches. Its configuration allows for increased flexibility and robustness in practical applications: indeed, by applying additional constraints for class assignment, it effectively identifies possible outliers. In this manner, Soft PLS-DA overcomes the limitations of PLS-DA in handling new objects not belonging to the considered classes, maximizing at the same time the discrimination between the classes of interest [42,49,50]. For these reasons, the soft discriminant approach represents a better fit compared to conventional discriminant analysis or class modelling strategies to face the problem at hand.

In Soft PLS-DA, model computation is analogous to conventional PLS-DA. First, a PLS2 model is built using the spectral matrix as predictor matrix (**X**) and a dummy response matrix (**Y**) encoding for class membership. More in detail, the **Y** matrix contains as many rows as the number of objects and as many columns as the number of classes. Class membership of each object is represented with binary coding, where a value of 1 indicates class membership while 0 indicates non-membership. Then, class assignment is performed considering a threshold value for each class based on the predicted y values and determined according to Bayes theorem.

In Soft PLS-DA, a new sample is assigned to a defined class according to the following criteria:

- Q residuals values falling within the 99.9% confidence limit of the model. This limit was set wide enough to consider as much as possible within classes variability, but allowing at the same time to exclude samples with a very poor fit to the model;
- y predicted values falling within an acceptability range for the considered class. The lower limit of this range is defined by the PLS-DA threshold value for the considered class, while the upper limit allows the rejection of objects located at the extremes of the Gaussian probability density function;
- the samples must be unambiguously assigned to one class only: this aspect is rather useful for multiclass classification issues since it solves ambiguities in class attribution.

The samples that do not match all the criteria defined by the Soft PLS-DA decision rules are not assigned to any class and automatically labelled as *Not Assigned* (NA). The reader is referred to Ref. [42] for a detailed description of Soft PLS-DA algorithm.

The classification models were built considering the whole spectral range and, secondly, taking into account the most relevant wavelengths for classification selected by sparse-based variable selection with sparse Soft PLS-DA (s-Soft PLS-DA).

s-Soft PLS-DA couples Soft PLS-DA algorithm described above with Least Absolute Shrinkage and Selection Operator (LASSO) penalisation, which constrains the sum of the absolute values of the regression vector to be lower than a tuning parameter. In this manner, s-Soft PLS-DA

algorithm simultaneously performs model calculation and variable selection by shrinking to zero the coefficients of irrelevant variables [51–55]. Therefore, the resulting sparse model relies exclusively on the variables with non-zero coefficients, while variables whose coefficients are equal to zero are effectively excluded from the classification model. This property improves model interpretability and allows the identification of the most informative spectral variables.

To obtain a good sparse solution, i.e., many regression coefficients shrunk to zero, it is necessary to choose the proper number of latent variables (LVs) and the level of sparsity, i.e., the number of spectral variables to be selected. To this aim, grid search was performed testing all the possible combinations between a number of LVs ranging from 1 to 10 and a number of spectral variables selected for each LV ranging from 5 to 100, with a step equal to 5.

Both Soft PLS-DA and s-Soft PLS-DA classification models were calculated considering the same preprocessing methods leading to the best results in the annotation procedure (Section 2.3.1), i.e., Multiplicative Scatter Correction, MSC (median spectrum as reference) and mean center for *cv*. Abate Fétel, and linear-detrend and mean center for *cv*. Williams.

The classification models were optimized using a custom cross-validation scheme, obtained by splitting the TR spectra based on pear sample into six deletion groups. For both cultivars, the optimal Soft PLS-DA and s-Soft PLS-DA models were selected as those leading to the highest Non-Error Rate (NER) values. Afterwards, the classification performances were assessed though external validation using the TS set of spectra.

The classification performances of the models were evaluated in calibration (CAL), cross-validation (CV) and prediction (PRED) of the external TS set by calculating the statistical parameters sensitivity (SENS), specificity (SPEC), efficiency (EFF), and Non-Error Rate (NER) [56].

2.4. From pixel-level classification models to image-level predictions

For each cultivar, the best performing Soft PLS-DA and s-Soft PLS-DA models were applied to the unfolded hyperspectral images, obtaining for each image a vector with as many elements as the pixels retained after background removal, and whose values correspond to the predicted class labels of each spectrum. These vectors of predicted class labels were then refolded to the original spatial dimensions of the corresponding hyperspectral images to generate the prediction images, i.e., grey-scale images where each pixel is assigned to *Sound* or *Punctured* class, or not assigned to any class (NA), according to the classification outcome of its corresponding spectrum. This allows to visualize the spatial distribution of the predicted classes over the fruit surface and have a qualitative assessment of model performances.

To simulate the application of the proposed approach to a fruit sorting line, it is necessary to move from the pixel-level prediction images to an image-level classification by classifying each image fruit as *Sound* or *Punctured*.

For the sake of clarity, from here onwards we will use the term *object* to refer to any cluster of connected pixels in a prediction image whose spectra are classified as belonging to the *Punctured* class by the selected Soft PLS-DA or s-Soft PLS-DA model.

For image-level classification, the prediction images were further elaborated by implementing an additional classification rule based on the size of the objects predicted as punctured areas. More in detail, each prediction image was converted into a binary mask with values equal to 1 for the pixels assigned to the *Punctured* class and values equal to 0 for all the other pixels (i.e., background pixels, pixels assigned to *Sound* class or not assigned pixels). To refine the irregular and jagged edges of punctured areas, morphological erosion was applied to each binary mask using a disk-shaped structuring element with radius equal to 2 pixels. Then, the *bwconncomp* function from the MATLAB Image Processing Toolbox was used to identify connected components in the binary masks, using an 8-pixels connectivity criterion. Based on this criterion, pixels were considered as belonging to the same connected component if they were connected along the horizontal, vertical or diagonal direction. The same function was also used to determine the number of objects in each image and the number of pixels included in each object.

The threshold value of 105 contiguous pixels was set as the minimum object size for punctures, as it corresponded to the minimum number of pixels of annotated ROIs of punctured areas in the TR images for both cultivars. To further verify this setting, several threshold values ranging from 90 to 200 contiguous pixels were tested by comparing the size of objects wrongly predicted as punctures with those correctly assigned to punctured regions. Table S4 in the Supplementary Material reports the image-level classification results for five tested threshold values applied to TR prediction images, expressed in terms of SENS, SPEC and EFF for the *Punctured* class for both cultivars. The threshold value of 105 contiguous pixels generally provided the best compromise in classification performance regardless of model type or pear cultivar, thus supporting the choice of this threshold value as the minimum object size for punctures.

Consequently, objects predicted as *Punctured* smaller than the established threshold were removed from the prediction images.

Based on this rule, images containing at least one object with more than 105 pixels predicted as *Punctured* were automatically categorized as "*Punctured images*", while images with no punctured objects or punctured objects smaller than 105 pixels were assigned to the "*Sound images*" category.

The image-level classification performances were quantitatively assessed considering the percentages of S–C images predicted as *Sound* and P-E annotated images predicted as *Punctured*.

Moreover, the same classification rule was applied to the prediction images of all the other fruit images to evaluate the behaviour of the models towards unknown damage, i.e., damage not caused by BMSB. The resulting prediction images were used to visualize the prediction performances directly on the original image domain in order to perform a global evaluation of the classification performances.

2.5. Software

Background removal and spectra extraction were performed on each image in an automated manner using routines written in MATLAB environment (R2020a, The MathWorks, USA), based on the Image Processing Toolbox (v. 11.1) and the PLS_Toolbox (v. 8.8.1, Eigenvector Research Inc., USA).

Soft PLS-DA models were calculated using the MATLAB function freely downloadable from https://www.chimslab.unimore.it/downloads/. s-Soft PLS-DA models and the final image-level classification step were implemented using *ad hoc* routines written in MATLAB environment.

3. Results and discussion

3.1. Classification at pixel-level

Table 3 reports the results obtained in CAL, CV and PRED of the external TS of representative spectra for the Soft PLS-DA and s-Soft PLS-DA models. For each model, the classification performances were evaluated considering SENS, SPEC, EFF, NA and NER expressed as percentage values for both *Punctured* and *Sound* classes.

Overall promising classification results were obtained in CV and PRED of the external TS set considering both cultivars. Starting from *cv*. Abate Fétel, the best performances in terms of NER were obtained retaining 6 LVs, both considering the entire spectrum or only the wavebands most relevant for discrimination purposes selected by sparse based methods. In the case of *cv*. Williams, the best performances were obtained by choosing 3 LVs for the calculation of Soft PLS-DA and 5 LVs

Table 3
Pixel-level classification results obtained by applying Soft PLS-DA and s-Soft PLS-DA in calibration (CAL), cross-validation (CV) and prediction (PRED) of the TS set. The results, expressed in terms of percentages of SENS, SPEC, EFF, and NA, are reported for both Punctured and Sound classes while NER provides a global evaluation of the classification performances.

			Abate Fétel			Williams		
			MSC (median)+ mc			Linear-detrend + mc		
			CAL	CV	PRED	CAL	CV	PRED
Soft PLSDA	LVs		6			3		
	Spectral variables		100			100		
	SENS (%)	Punctured	91.9	90.4	**93.3**	93.8	92.1	**91.8**
		Sound	94.2	93.8	**88.0**	95.2	95.6	**97.2**
	SPEC (%)	Punctured	96.8	96.5	**90.8**	95.7	96.1	**97.4**
		Sound	94.5	93.7	**95.6**	94.1	92.9	**91.8**
	EFF (%)	Punctured	94.3	93.4	**92.0**	94.7	94.1	**94.5**
		Sound	94.4	93.7	**91.7**	94.6	94.2	**94.5**
	NA (%)	Punctured	2.7	3.2	**2.1**	0.3	0.8	**0.1**
		Sound	2.6	2.7	**2.6**	0.5	0.5	**0.2**
	NER		93.0	92.1	**90.7**	94.5	93.8	**94.5**
s-Soft PLSDA	LVs		6			5		
	Spectral variables		**51**			**54**		
	SENS (%)	Punctured	90.5	89.1	**91.8**	93.0	92.8	**93.0**
		Sound	95.8	95.4	**91.2**	94.6	95.7	**89.3**
	SPEC (%)	Punctured	96.3	96.1	**91.9**	95.4	97.1	**91.6**
		Sound	92.4	91.6	**93.4**	94.3	95.2	**94.2**
	EFF (%)	Punctured	93.4	92.6	**91.9**	94.2	94.9	**92.3**
		Sound	94.1	93.5	**92.3**	94.5	95.5	**91.7**
	NA (%)	Punctured	1.9	2.5	**1.3**	1.3	2.4	**1.1**
		Sound	0.5	0.7	**0.8**	0.8	1.4	**2.3**
	NER (%)		93.1	92.3	**91.5**	93.8	94.3	**91.2**

for the calculation of the sparse model.

Despite the objective of the work is to correctly detect *Punctured* areas, the least possible amount of *Sound* areas must be misclassified to avoid waste and economical losses. In this context, NER could be a suitable indicator to assess the overall ability of the model to correctly assign both *Punctured* and *Sound* spectra to the class they belong to. In prediction of the external TS set (PRED), the models achieving the best results for both cultivars led to NER values greater than 90.7% at the pixel-level, which is promising due to the complexity of the problem at hand.

The number of spectra rejected as NA is mostly higher for *cv*. Abate Fétel, in particular for the Soft PLS-DA model. Indeed, for *cv*. Abate Fétel, the percentage of spectra predicted as NA decreases for the s-Soft PLS-DA model for both classes investigated. Generally, s-Soft PLS-DA leads to slightly better results in PRED, reaching a NER of 91.5% at pixel-level. Conversely, considering *cv*. Williams, an increase of the percentage of spectra predicted as NA is shown for the s-Soft PLS-DA model. Although s-Soft PLS-DA exhibited a minor reduction in PRED performance, especially for the *Sound* class, it still represents a valid solution, achieving a NER of 91.2% while considering only half of the original spectral variables.

Generally, it has to be highlighted that s-Soft PLS-DA models led to classification results comparable to those of Soft PLS-DA, but proposing a more parsimonious solution in terms of considered spectral variables. Indeed, as reported in Table 3, s-Soft PLS-DA allowed to reduce the original 100 spectral variables by selecting only 51 and 54 variables for *cv*. Abate Fétel and *cv*. Williams, respectively.

Fig. 2 reports the regression vectors for the *Punctured* class of the models calculated considering the whole spectrum (dotted line) or the sparse solution (solid line), enabling a possible interpretation of the spectral regions that differentiate *Punctured* from *Sound* areas of the fruits. Notably, the spectral variables selected by the two s-Soft PLS-DA models, derived from the spectra of the two pear cultivars, are largely consistent. The spectral regions selected by both sparse models are highlighted in green colour in Fig. 2, and only these spectral regions contributed to the calculation of the s-Soft PLS-DA models.

The main spectral regions with high absolute values for both varieties fall into the 1150−1180 nm (C═O, C–H second overtone), in the

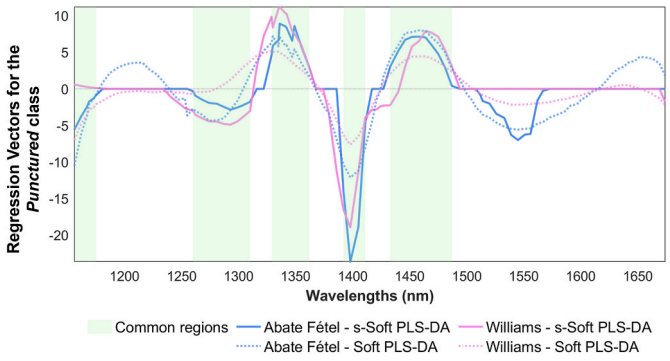


Fig. 2. Regression vectors for cv. Abate Fétel and cv. Williams: the dotted lines refer to the Soft PLS-DA regression vectors while the solid lines refer to the regression vectors obtained with sparse based variable selection. The areas highlighted in green refer to common spectral regions selected for both cultivars. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

1260−1310 nm (C–H second overtone), in the 1330−1370 nm (C–H_3 combination band), in the 1390-1420 nm (C–H_2 combination band, aromatic C–H combination band and O–H stretch first overtone), and in the 1430−1490 nm (O–H first overtone, C═O stretch third overtone and N–H stretch first overtone) spectral regions. Moreover, the region selected only for *cv*. Abate Fétel falling in the interval at 1510-1570 nm may be related to starch content, which is gradually broken down into soluble sugars during fruit ripening. Conversely, for *cv*. Williams there are not actual additional spectral regions considered, just isolated wavebands that slightly expand the intervals already selected [57–59].

The spectral regions falling in the intervals at 1330–1370 nm and at 1430–1490 nm could be associated with a higher absorption of bonded water, carbohydrates and fibres which characterize the *Sound* areas of the fruits. Conversely, the spectral bands falling in the intervals at 1150−1180 nm, 1260−1310 nm and 1390-1420 nm may refer to a higher absorption of free water, sugars and phenolic compounds in *Punctured* areas. Interestingly, the spectral regions selected seem coherent with the most informative spectral regions identified in the previous study related to the annotation of BMSB punctures (*see* Ref. [39] Supplementary Material).

The suggested chemical interpretation of the selected spectral regions provided above is supported by several studies that reported changes in the sugars and polyphenols in fruits punctured by BMSB as a metabolic response [60–63]. Interestingly, Zamljen and colleagues [63] reported that the metabolic response of apples fruits (*Malus* spp.) towards BMSB punctures involves an increase in sugars and polyphenols (i.e., hydroxycinnamic acids, flavanols) content only in the damaged areas. To our knowledge, similar studies have not been performed on pears. However, both apple (*Malus* spp.) and pear (*Pyrus* spp.) fruits may exhibit a similar metabolic behaviour since they belong to the *Maloideae* sub-family [64,65].

3.2. Classification at image-level

As mentioned in Section 2.4, the prediction images obtained by applying the classification models to the hyperspectral images were further elaborated to obtain an image-level prediction. In particular, only those prediction images containing at least one *Punctured* object larger than 105 pixels were classified as *Punctured images*, while the other images were considered as *Sound images*. This object-based classification rule was necessary to translate the prediction images into an image-level decision. Indeed, the prediction maps showed that, in addition to the actual punctured regions, also some isolated pixels of sound tissue in P-E images and of sound fruits in S–C images were occasionally assigned to the *Punctured* class. This outcome is very common when applying classification models to whole hyperspectral images, due to increased variability sources that cannot be fully represented during model calculation, such as shape irregularities, lighting effects or peel russeting, among others. These isolated misclassified pixels can be regarded as an intrinsic error of the classification models, while actual punctures appeared as spatially connected clusters of pixels predicted as *Punctured*. The threshold of 105 pixels to identify objects predicted as punctures was thus necessary to limit as much as possible the false positives, i.e., images of sound fruits wrongly classified as *Punctured* fruits. Of course, the selected threshold represents a compromise between minimising false positive detections caused by isolated misclassified pixels and maintaining a good ability of detecting small punctured regions.

The threshold of 105 connected pixels was established based on the smallest annotated ROI corresponding to punctured tissue in TR images (see Section 2.4). Since this value was identified across images acquired at all time points, the selected threshold inherently accounts for potential variations in injury size associated with the progression of damage over time. However, it should be noted that the increase in puncture size during storage was relatively limited when compared with the overall variability in damage size observed among fruits.

To provide a practical interpretation of the prosed threshold value, it is possible to relate the minimum cluster size of 105 connected pixels to its corresponding physical area, as this value represents the smallest punctured region that can be reliably identified using the proposed approach. Given the spatial resolution of the hyperspectral system of 0.2 mm, a region of 105 connected pixels corresponds to an area of approximately 4.2 mm^2 (=105 × 0.2 mm × 0.2 mm), which is equivalent to a circular lesion with a diameter of about 2.3 mm.

The visual observation of damage size performed after peeling the fruits was consistent with the findings in Ref. [66], where an average internal injury area of 15 ± 6 mm^2 was reported for apples exposed to BMSB for 48 h in July and evaluated in August. Therefore, the smallest punctured region that can be reliably identified using the proposed approach is sufficient to detect the injured spots caused by BMSB punctures, including lesions smaller than the average damage size reported in the literature, still visible to the naked eye.

The threshold value was defined by jointly considering both cultivars and both classification approaches; however, in specific applications it could be further optimized according to cultivar, classification model and characteristics of the spectral acquisition system. However, these aspects are out of the aim of current study.

Table 4 reports the results obtained by applying Soft PLS-DA and s-Soft PLS-DA models to the TR (CAL) and TS (PRED) hyperspectral images, which belong to S–C and annotated P-E categories. The results are expressed as percentages of images predicted as *Punctured* or predicted as *Sound*, concerning both P-E annotated and S–C images individually.

Comparing the results obtained from the two cultivars, better global performances were obtained for the *cv*. Williams compared to *cv*. Abate Fétel.

In general, the ability of the models in correctly identifying the P-E images as *Punctured* are generally higher than the ability in recognizing the S–C images as *Sound*. This is aspect is particularly critical for the *cv*. Abate Fétel classification model considering the whole spectrum, that allowed to recognize as *Sound* 89.8% and 63.3% of the S–C images belonging to the TR set and external TS set, respectively. An improvement was achieved considering only the spectral regions selected by sparse based methods (s-Soft PLS-DA), enabling the correct assignment of 90.8% and 74.3% of the S–C images belonging to the TR set and external TS set, respectively. This behaviour may be attributed to the higher degree of peel russeting and the more irregular fruit shape of *cv*. Abate Fétel compared with *cv*. Williams, which likely increased the spectral variability of sound tissues and, consequently, the probability of misclassification. In contrast, the models' ability in recognizing correctly the P-E annotated images as *Punctured images* is definitely satisfactory, achieving 81.5% and 82.7% in prediction of the external TS

Table 4
Image-level classification results of Soft PLS-DA and s-Soft PLS-DA models. For each cultivar, the prediction results for the P-E annotated and the S–C images, subdivided in TR and TS images, are reported. SENS values, i.e., P-E annotated images correctly predicted as Punctured images, are coloured in red while SPEC values, i.e., S–C images correctly rejected by the Punctured class, are coloured in green.

			Abate Fétel			Williams		
			# images	Predicted as *Punctured* (%)	Predicted as *Sound* (%)	# images	Predicted as *Punctured* (%)	Predicted as *Sound* (%)
Soft PLSDA	P-E Annotated	TR	**129**	**87.6**	12.4	**182**	**97.3**	2.7
		TS	**81**	**81.5**	18.5	**91**	**94.5**	5.5
	S-C	TR	**412**	10.2	**89.8**	**340**	10.3	**89.7**
		TS	**237**	36.7	**63.3**	**240**	5.4	**94.6**
s-Soft PLSDA	P-E Annotated	TR	**129**	**86.8**	13.2	**182**	**97.3**	2.7
		TS	**81**	**82.7**	17.3	**91**	**96.7**	3.3
	S-C	TR	**412**	9.2	**90.8**	**340**	15.6	**84.4**
		TS	**237**	25.7	**74.3**	**240**	13.3	**86.7**

set images considering the whole spectrum and sparse variable selection, respectively.

On the other hand, the classification results for *cv*. Williams were far more promising. Considering the whole spectrum, the ability of *cv*. Williams Soft PLS-DA model outperformed the *cv*. Abate Fétel one, leading to the correct assignment of 94.5% of the P-E annotated images and 94.6% of the S–C images belonging to the external TS set. Throughout sparse based variable selection, the classification results were slightly worse, leading to the correct assignment of 96.7% of the P-E annotated images and of 86.7% of the S–C images belonging to the external TS set.

In Figs. 3 and 4 some representative prediction images for both cultivars are reported along with the corresponding RGB images of peeled and unpeeled fruits used as reference. For *cv*. Abate Fétel, the prediction images of S–C and P-E images belonging to the TS are shown in Fig. 3A–C.

As a result of the image level decision rule, the S–C image in Fig. 3 B is wrongly assigned to the *Punctured* class by Soft PLS-DA, by predicting a rusty superficial spot on the fruit peel as puncture. Conversely, s-Soft PLS-DA is able to correctly predict the S–C image as *Sound*, and also reduce the number of NA pixels. Both S–C and P-E images shown in Fig. 3 A and 3C are correctly predicted by the models. Nonetheless, the

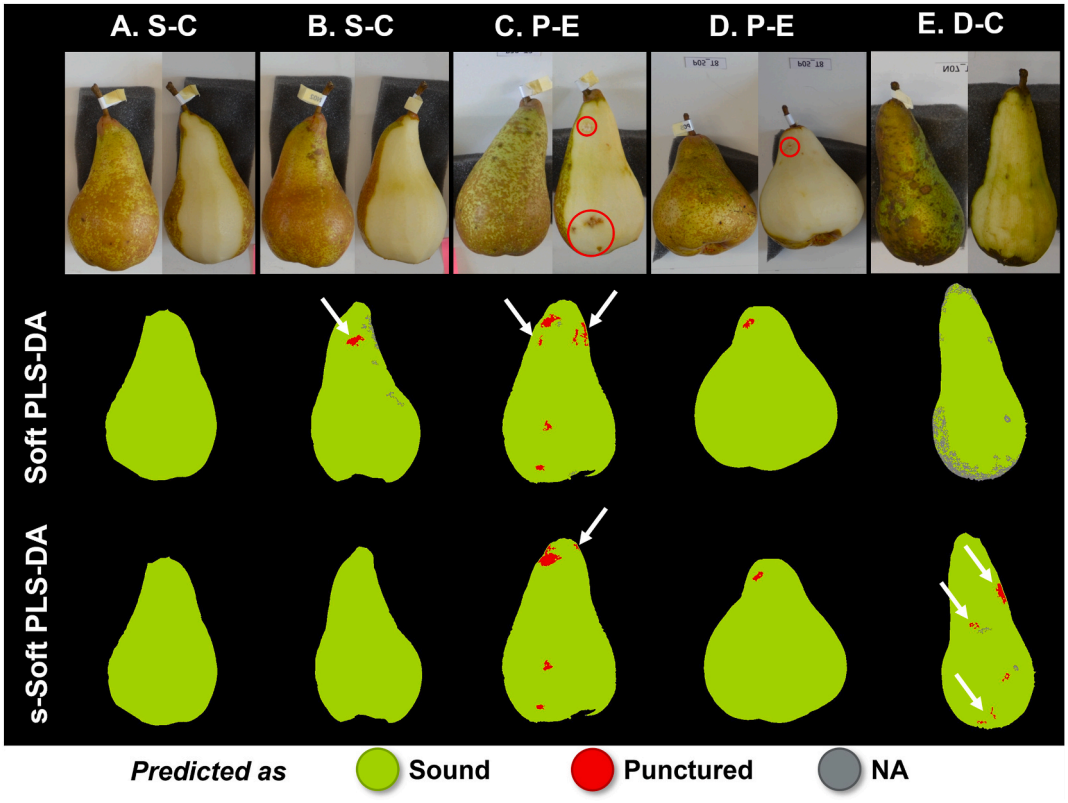


Fig. 3. *cv*. Abate Fétel prediction images obtained by applying Soft PLS-DA and s-Soft PLS-DA models together with the corresponding RGB images of the samples.

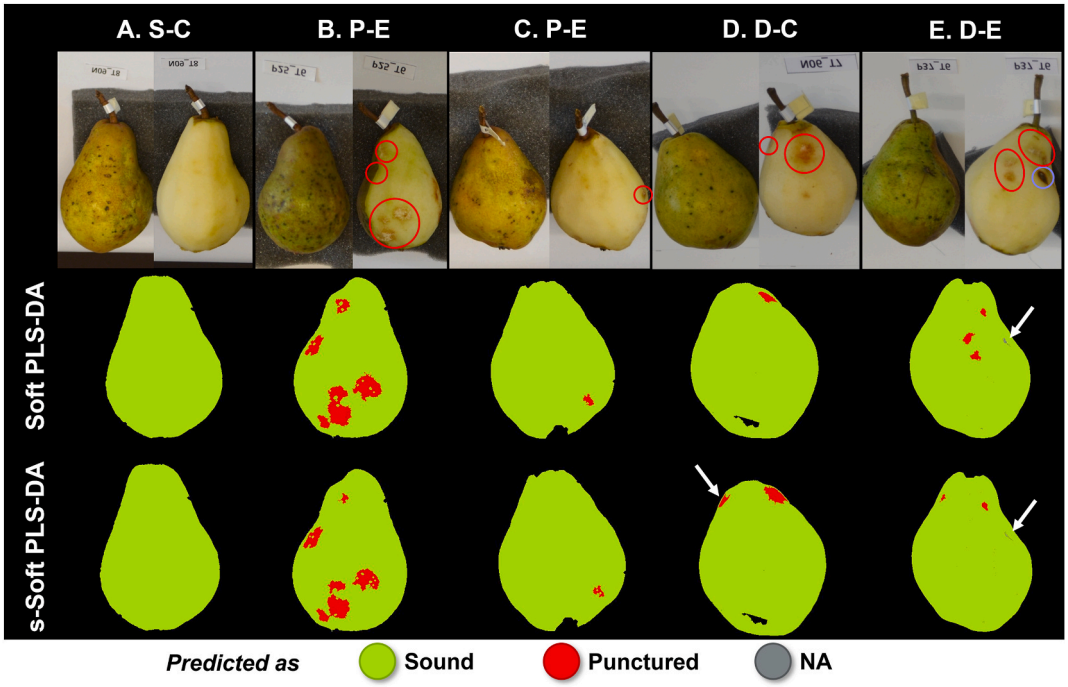


Fig. 4. *cv.* Williams prediction images obtained by applying Soft PLS-DA and s-Soft PLS-DA models together with the corresponding RGB images of the samples.

prediction image obtained applying Soft PLS-DA on the P-E image shows some additional areas wrongly predicted as *Punctured*, whose extension is reduced when considering s-Soft PLS-DA results.

Fig. 3 D shows the prediction images of a P-E image whose punctured areas were not previously annotated by the methodology followed in Ref. [39]. In this case, despite being quite small, the area ascribable to the puncture was correctly seen by both Soft PLS-DA and s-Soft PLS-DA. In Fig. 3 E an image of a Damaged-Control (D-C) fruit section is shown as an example of how the models may behave towards other types of damage. In this case, the pear is affected by pear scab (i.e., other type of damage, not ascribable to the BMSB): Soft PLS-DA effectively predicts several areas of the fruit as NA, while the sparse solution assigns them as *Punctured* areas. Despite handling this situation in a different manner, both models seem to partially identify areas of the fruit that are not compliant to the *Sound* class.

The same evaluation can be done for *cv.* Williams prediction images shown in Fig. 4. In this case, S–C and P-E images in Fig. 4A and B belong to the TS set: both images were correctly predicted as *Sound* and *Punctured images* by Soft PLS-DA and s-Soft PLS-DA, with no significant differences. As for *cv.* Abate Fétel, the prediction images of a P-E image whose punctured areas were not previously annotated is shown in Fig. 4C. Throughout the comparison with RGB reference images, both Soft PLS-DA and s-Soft PLS-DA models enabled the identification of punctured area.

Nonetheless, suberification and necrosis of the fruit pulp are a common response of plants and corresponding fruits to stress, that may be due to different causes. For *cv.* Williams, an additional evaluation of models' robustness was done towards D-C images showing suberification damage (Fig. 4 D). Regardless of the agents that caused them to emerge, suberified tissues have the same chemical composition therefore compromising fruits' quality. As for the annotation methodology in Ref. [39], the calculated classification models are capable to detect damage sharing the same chemical composition as *Punctured* areas, although not directly being caused by BMSB.

Another interesting scenario is shown in Fig. 4 E, in which a fruit exposed to BMSB exhibits both BMSB-induced suberification and a moulded area. The corresponding prediction images calculated applying Soft PLS-DA and s-Soft PLS-DA partially detect the suberified areas, correctly predicting the image as *Punctured*. Coherently with Fig. 3 E, part of the moulded area is predicted as NA by both models, suggesting that different damage not previously considered for model training may be present.

The ability of the classification models towards the identification of punctures during post-harvest storage is consistent, despite the major impact of the different fruits' rotation during the subsequent image acquisition times [39].

Figs. 5 and 6 display the predictions for each P-E and S–C section over time for *cv.* Abate Fétel and *cv.* Williams, respectively. The predictions are presented in a heatmap-style layout, where each column represents a specific fruit section and the rows correspond to the different acquisition times. Accordingly, each cell represents the image-level classification of a single hyperspectral image: red cells indicate images classified as *Punctured*, green cells correspond to images classified as *Sound*, and grey cells indicate that no image was acquired for that section at the corresponding acquisition time.

Overall, punctures seem detectable from the harvest day (T1), becoming slightly more defined around the edges over time. Concerning both cultivars, the P-E sections predicted as *Punctured* at least once by Soft PLS-DA and s-Soft PLS-DA are basically the same. On the other hand, the number of times that P-E sections are predicted as *Punctured* is slightly different depending on the model, since s-Soft PLS-DA improves the predictions for the *Punctured* class.

The same conclusion cannot be drawn for the S–C images. Concerning *cv.* Abate Fétel, s-Soft PLS-DA model seem to misclassify more often S–C sections acquired at T1 (day of harvest): interestingly, the same behaviour is shown for the prediction of S-E sections (Fig. S1 of Supplementary Material). Conversely, this trend is not shown for Soft PLS-DA, where the wrong assignations are repeated for the same S–C sections over time, probably due to superficial spots on the fruit peel (e.g., section C of sample N02 harvested in 2022, also reported in Fig. 3 B). This behaviour may be partially explained by the relevance of the spectral region at 1510-1570 nm for *cv.* Abate Fétel s-Soft PLS-DA model. Interestingly, this spectral region seems related to starch content [57], compound that gradually decrease during fruit ripening since being converted in soluble sugars (*see* Fig. 2).

For *cv.* Williams, while the predictions for S–C sections obtained with

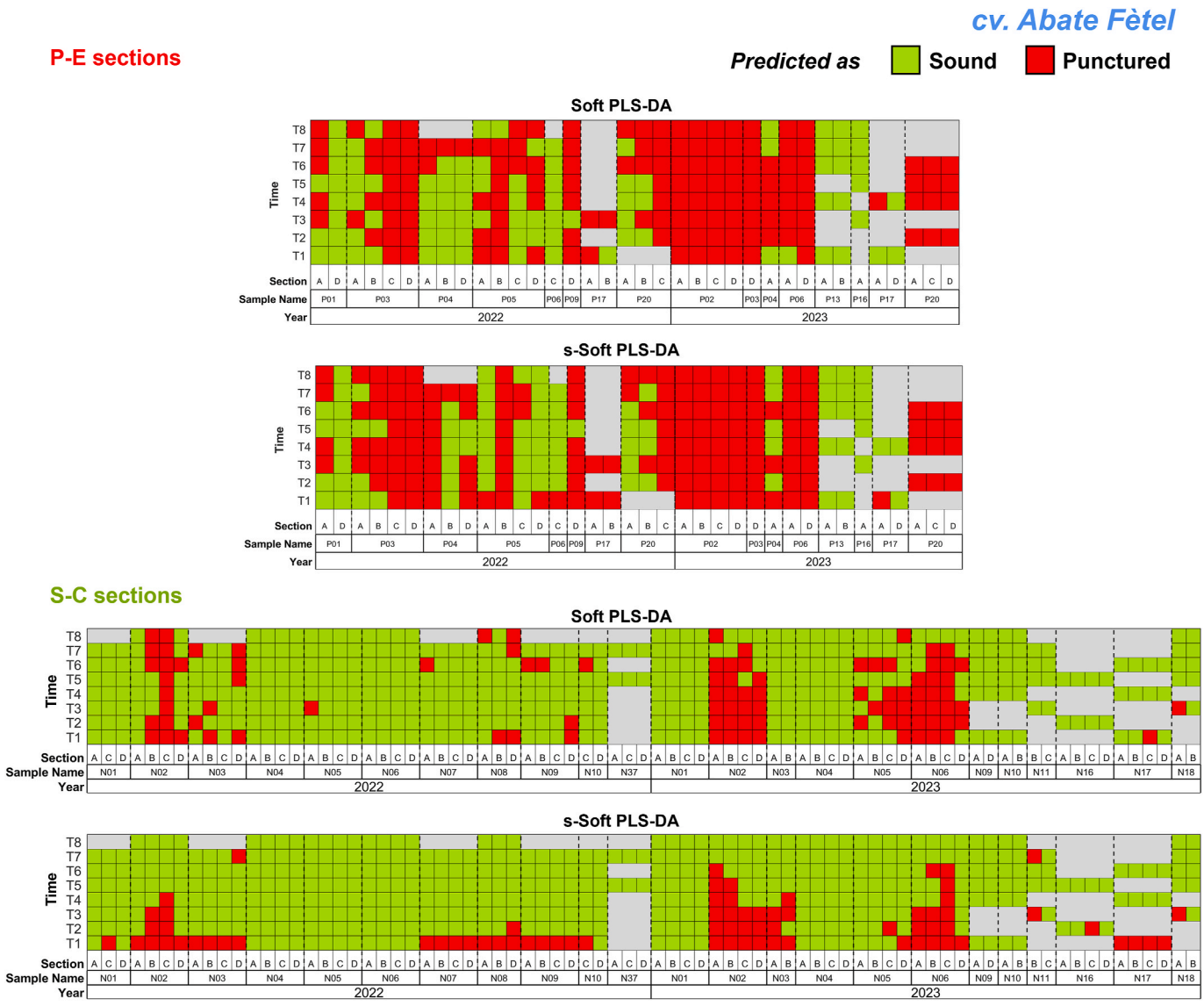


Fig. 5. cv. Abate Fétel: image-level predictions for P-E and S–C sections over time, considering Soft PLS-DA and s-Soft PLS-DA models.

Soft PLS-DA do not show significant differences between vintages, s-Soft PLS-DA tends to misclassify more often S–C sections of fruits harvested in 2022. A similar behaviour is shown in predictions for the S-E sections (Fig. S2 of Supplementary Material), where the sections predicted as *Punctured* at least one time are more or less the same for fruits harvested in 2023 while the frequency of misclassifications is higher for fruits harvested in 2022.

Table 5 reports the global classification performances at image level towards modelled classes (i.e., P-E and S–C images) and unknown classes (i.e., D-C, D-E, S-E images), in terms of percentages of images seen as *Punctured* or seen as *Sound* by the models.

The D-C and D-E images showing mild damage (Type 1) of both cultivars are generally seen as *Sound* by the models, and this can be explained considering the low severity of the damage. On the other hand, for the D-C images showing severe damage not due to BMSB (Type 3) the percentages of images seen as *Punctured* by the models tend to increase. In this case, this category comprehends all the severe damage determined by abiotic or biotic agents such as moulds, scab but also suberifications or pulp necrosis not due to BMSB. As a matter of fact, the percentages of D-C (Type 3) images seen as *Punctured* exceed the 60% for *cv.* Williams, for which the occurrence of suberified areas on fruits is a common plant response to adversities (Fig. 4 D).

The D-E images showing severe damage not ascribable to BMSB (Type 3) are generally affected by moulds, scab or chilling injuries. These injuries are mainly seen as *Sound* by the models, which may be due to the different chemical composition from suberified areas. A particular case is depicted in Fig. 4 E, in which a D-E image is showing a punctured and a moulded area: the two images falling into this scenario are seen as *Punctured*. Analogous considerations can be made comparing the predictions for P-E sections and other damaged sections (i.e., D-C, D-E) reported in Figs S3 and S4 of Supplementary Material.

Concerning the ability of the models to correctly identify *Sound* pears, the percentages of images labelled as S–C and S-E predicted as *Sound* can be compared. The worst results were obtained for *cv.* Williams when using s-Soft PLS-DA, where the percentage of S-E images predicted as *Sound* was up 21% less than S–C images. The lower classification performances seem emphasized for S-E images of fruits harvested in 2022 (*see* Table S6 of Supplementary Material). Interestingly, the worst outcomes coincide with *cv.* Williams fruits harvested in 2022, which were deeply affected by extreme high temperatures registered at the beginning of August 2022. The adverse climatic conditions may had a negative impact, not only for BMSB specimens' vitality in the field and the amount of P-E images collected [39], but on the overall quality of the harvested fruits.

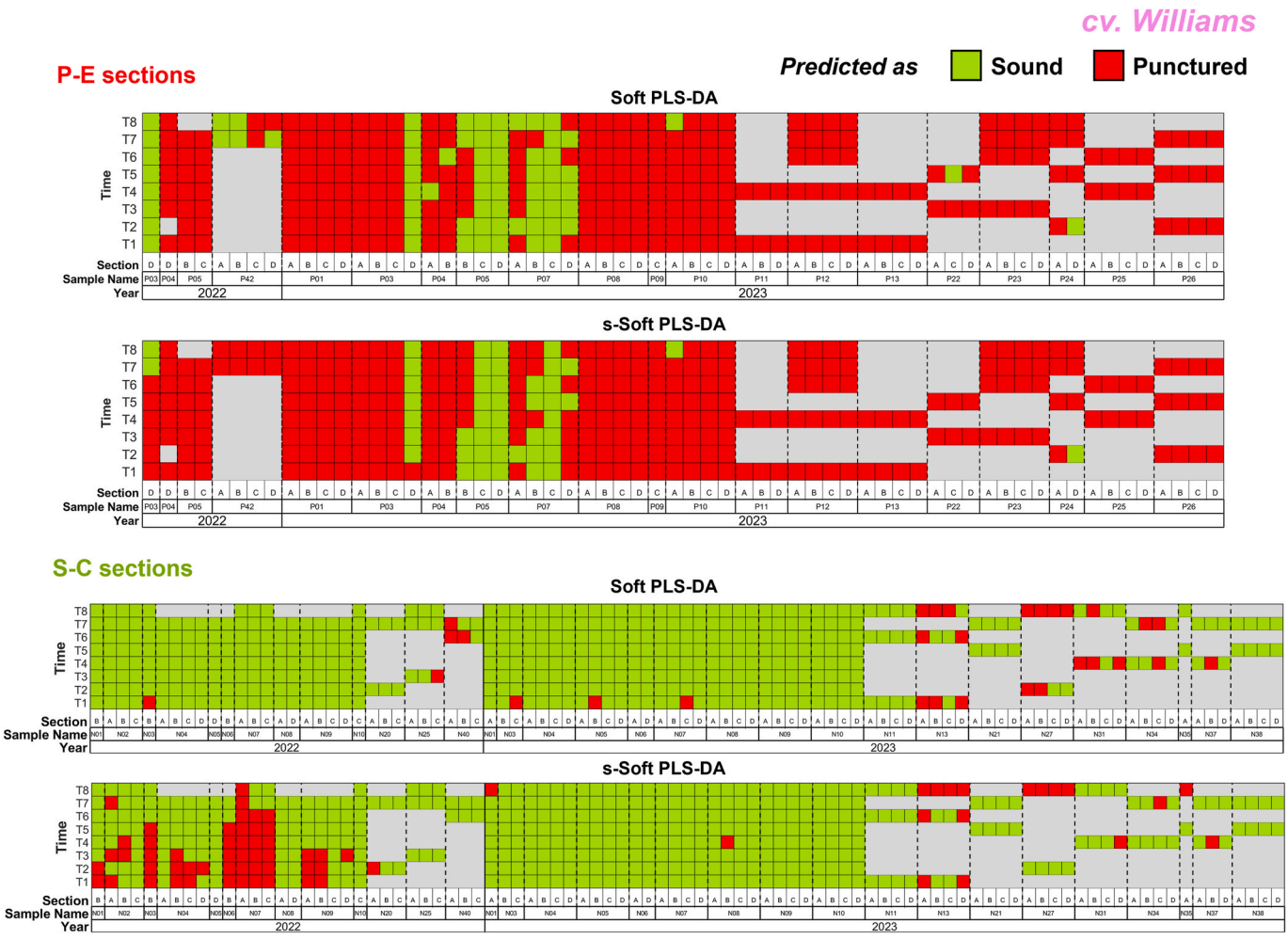


Fig. 6. cv. Williams: image-level predictions for P-E and S–C sections over time, considering Soft PLS-DA and s-Soft PLS-DA models.

Conversely, the results for *cv*. Abate Fétel are more promising: indeed, the percentage of images predicted as *Sound* of S–C and S-E are comparable for both Soft PLS-DA and s-Soft PLS-DA.

4. Conclusions

The present study builds upon our previous work [39] to present a complete workflow for detecting BMSB-induced punctures in pears using NIR-HSI, from the annotation of punctured regions to the development of classification models, providing relevant insights for the future development of post-harvest sorting systems.

Importantly, the fruit samples were exposed to BMSB under real field conditions over two harvest years, capturing the natural variability of damage and leading to the acquisition and elaboration of a large number of hyperspectral images.

More in detail, this study aimed at developing robust and flexible classification models able to identify damage due to BMSB punctures on pears. As previously discussed [39], the identification of regions ascribable to BMSB punctures is a challenging task due to the complexity of the dataset and the slight chemical differences between sound and damaged areas. However, a preliminary step allowed the annotation of a consistent number of regions ascribable to punctures, thus supporting the feasibility of a more detailed pixel-level classification.

To enable classification, representative spectra of sound and punctured areas were randomly sampled from sound sections of control fruits (S–C) and images of punctured sections of fruits exposed to BMSB (P-E), respectively. The classification models were calculated considering Soft PLS-DA algorithm, which provides a flexible and robust solution in sorting systems since it is able to handle unmodelled classes. In addition, Soft PLS-DA was combined with sparse based variable selection (s-Soft PLS-DA) to identify the most important spectral variables for classification. Both Soft PLS-DA and s-Soft PLS-DA showed promising results at pixel-level, achieving prediction efficiencies of at least 91.9% for *cv*. Abate Fétel and 92.3% for *cv*. Williams for the *Punctured* class.

Furthermore, an additional threshold on the size of objects predicted as *Punctures* was set enabling image-level classification. Image-level classification is more suitable for online sorting systems, since it allows the automated assignment of the samples as a whole to a certain class. For *cv*. Williams, comparable classification performances were achieved at both pixel and image levels, with classification efficiencies in prediction of at least 92.3% and 91.6%, respectively. Conversely, image-level performance for *cv*. Abate Fétel was lower – but still satisfactory – leading to 71.8% and 78.4% classification efficiencies in prediction for Soft PLS-DA and s-Soft PLS-DA, respectively.

These results likely reflect the heterogeneity of defects, subtle chemical differences between punctures and sound areas, and inherent variability within the images and between the fruits.

Beyond the objectives of the HALY.ID EU project for which the investigation was followed, training the models to detect other defects found on the studied fruits could improve the classification efficiency in post-harvest sorting systems.

Nevertheless, encouraging outcomes were obtained when considering only the relevant wavebands for classification selected by s-Soft PLS-DA. Importantly, most spectral regions relevant for discrimination

Table 5
Global evaluation of image-level predictions of Soft PLS-DA and s-Soft PLS-DA models for both cv. Abate Fétel and cv. Williams.

			Soft PLSDA		s-Soft PLSDA	
		# images	Predicted as *Punctured* (%)	Predicted as *Sound* (%)	Predicted as *Punctured* (%)	Predicted as *Sound* (%)
Abate Fétel	D-C (Type 1)	122	14.8%	85.2%	20.5%	79.5%
	D-C (Type 3)	12	41.7%	58.3%	50.0%	50.0%
	D-E (Type 1)	247	37.7%	62.3%	34.4%	65.6%
	D-E (Type 3)	63	41.3%	58.7%	39.7%	60.3%
	P-E	322	67.4%	32.6%	68.6%	31.4%
	S-C	653	19.9%	80.1%	15.6%	84.4%
	S-E	260	13.5%	86.5%	10.4%	89.6%
Williams	D-C (Type 1)	207	18.8%	81.2%	28.5%	71.5%
	D-C (Type 3)	152	61.8%	38.2%	68.4%	31.6%
	D-E (Type 1)	263	27.8%	72.2%	39.5%	60.5%
	D-E (Type 3)	74	43.2%	56.8%	23.0%	77.0%
	P-E	352	78.7%	21.3%	84.1%	15.9%
	S-C	582	8.2%	91.8%	14.6%	85.4%
	S-E	331	25.1%	74.9%	35.6%	64.4%

were consistent across both cultivars and appear to be associated with slight compositional differences involving sugars and polyphenols contents, in agreement with literature [59–62].

The selection by s-Soft PLS-DA of a limited number of spectral regions is particularly relevant since it provides a basis for the possible implementation of multispectral imaging (MSI) systems, which are more suitable for industrial sorting applications due to lower costs, increased acquisition speed, higher robustness of the optical components, and reduced data complexity.

A preliminary assessment of the feasibility of a MSI system could be carried out using the collected hyperspectral data to simulate a multispectral imaging system embedding only band-pass filters falling in the selected spectral regions [40].

However, the actual implementation of MSI systems for industrial sorting applications requires further investigation and validation under realistic operating conditions. In particular, future studies should evaluate the effects of conveyor-belt speed, image acquisition settings, illumination variability, and computational demand for data elaboration, in order to find the optimal operational conditions compatible with fruit sorting systems.

A critical aspect in the proposed application consists in the fact that damaged spots are spatially located in a specific fruit area, with the need of acquiring multiple portions of the same fruit to cover its whole surface. This problem can be solved using rotating roller conveyors and multi-view imaging solutions, that are commonly incorporated in modern fruit sorting systems.

Finally, although the models developed in this work used data collected over two harvest years, further studies should specifically investigate their robustness across seasons and environmental conditions. The inclusion of additional harvest years, orchards, and climatic conditions, together with model updating and recalibration, may contribute to improving long-term model generalisation and operational reliability.

CRediT authorship contribution statement

Veronica Ferrari: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Rosalba Calvini:** Conceptualization, Data curation, Investigation, Methodology, Software, Visualization, Writing – review & editing. **Camilla Menozzi:** Data curation, Investigation, Writing – review & editing. **Elena Costi:** Investigation, Resources. **Peter Offermans:** Funding acquisition, Project administration, Resources. **Lara Maistrello:** Funding acquisition, Project administration, Resources. **Alessandro Ulrici:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Authors wish to thank HALY.ID project of ERA-NET Cofund ICT-AGRI-FOOD, with funding provided by national sources (Ministero delle politiche agricole e forestali, MIPAAF) and co-funding by the European Union's Horizon 2020 research and innovation program, Grant Agreement number 862671.

Dr. Rosalba Calvini would like to thank the Italian funding programme Fondo Sociale Europeo REACT-EU - PON "Ricerca e Innovazione" 2014 – 2020 – Azione IV.6 Contratti di ricerca su tematiche Green (D.M. 1062 del 10/08/2021) for supporting her research (CUP: E95F21002330001; contract number 17-G-13884–4).

The authors wish also to acknowledge Dr. Daniele Giannetti and Dr. Niccolò Patelli from the Applied Entomology Lab (UNIMORE) for the support on-field.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2026.105817.

Data availability

Data will be made available on request.

References

[1] C.J.A. Bradshaw, B. Leroy, C. Bellard, D. Roiz, C. Albert, A. Fournier, M. Barbet-Massin, J.-M. Salles, F. Simard, F. Courchamp, Massive yet grossly underestimated global costs of invasive insects, Nat. Commun. 7 (2016) 12986, https://doi.org/10.1038/ncomms12986.

[2] IPBES, Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, [Object Object], 2019, https://doi.org/10.5281/ZENODO.3831673.

[3] T.C. Leskey, A.L. Nielsen, Impact of the invasive brown marmorated stink bug in North America and Europe: history, biology, ecology, and management, Annu. Rev. Entomol. 63 (2018) 599–618, https://doi.org/10.1146/annurev-ento-020117-043226.

[4] L. Maistrello, Case study 2: *halyomorpha halys* (Stål) in Europe, in: A.F. Bueno, A. R. Panizzi (Eds.), Stink Bugs Hemiptera Pentatomidae Res. Manag. Recent Adv. Case Stud. Braz. Eur. USA, Springer Nature Switzerland, Cham, 2024, pp. 271–359, https://doi.org/10.1007/978-3-031-69742-5_15.

[5] K.B. Rice, C.J. Bergh, E.J. Bergmann, D.J. Biddinger, C. Dieckhoff, G. Dively, H. Fraser, T. Gariepy, G. Hamilton, T. Haye, A. Herbert, K. Hoelmer, C.R. Hooks, A. Jones, G. Krawczyk, T. Kuhar, H. Martinson, W. Mitchell, A.L. Nielsen, D. G. Pfeiffer, M.J. Raupp, C. Rodriguez-Saona, P. Shearer, P. Shrewsbury, P. D. Venugopal, J. Whalen, N.G. Wiman, T.C. Leskey, J.F. Tooker, Biology, ecology, and management of brown marmorated stink bug (Hemiptera: pentatomidae), J. Integr, Pest Manag. 5 (2014) 1–13, https://doi.org/10.1603/IPM14002.

[6] A.L. Acebes-Doria, T.C. Leskey, J.C. Bergh, Injury to apples and peaches at harvest from feeding by *Halyomorpha halys* (Stål) (Hemiptera: pentatomidae) nymphs early and late in the season, Crop Prot. 89 (2016) 58–65, https://doi.org/10.1016/j.cropro.2016.06.022.

[7] M. Bariselli, R. Bugiani, L. Maistrello, Distribution and damage caused by *Halyomorpha* halys in Italy, EPPO Bull. 46 (2016) 332–334, https://doi.org/10.1111/epp.12289.

[8] A.L. Nielsen, G.C. Hamilton, Seasonal occurrence and impact of *Halyomorpha halys* (Hemiptera: pentatomidae) in tree fruit, J. Econ. Entomol. 102 (2009) 1133–1140, https://doi.org/10.1603/029.102.0335.

[9] E. Costi, T. Haye, L. Maistrello, Biological parameters of the invasive brown marmorated stink bug, *Halyomorpha halys*, in southern Europe, J. Pest. Sci. 90 (2017) 1059–1067, https://doi.org/10.1007/s10340-017-0899-z.

[10] CSO Italy Centro Servizi Ortofrutticoli, Estimation of damage from brown marmorated stink bug and plant pathologies related to climate change, CSO Italy Cent. Serv. Ortofrutticoli (2020). https://www.csoservizi.com/. (Accessed 28 October 2025).

[11] FAOSTAT, Food and Agriculture Organization of the United Nations Statistics Division, Crops and Livestock Products, Available online, 2025. https://www.fao.org/faostat/en/#data/QCL/visualize. (Accessed 28 October 2025).

[12] ISTAT, Italian National Institute of Statistics, 2025. https://esploradati.istat.it/databrowser/. (Accessed 28 October 2025).

[13] European Union, Regulation (EU) no. 1151/2012 of the European parliament and of the council of 21 November 2012 on quality schemes for agricultural products and foodstuffs. https://eur-lex.europa.eu/eli/reg/2012/1151/oj/eng, 2012. (Accessed 28 October 2025).

[14] S. Musacchi, I. Iglesias, D. Neri, Training systems and sustainable orchard management for European pear (Pyrus communis L.) in the mediterranean area: a review, Agronomy 11 (2021) 1765, https://doi.org/10.3390/agronomy11091765.

[15] Pera dell'Emilia-Romagna PGI Consortium, Production Regulations for the "*Pera dell'Emilia-Romagna*, PGI, 2023. https://peradellemiliaromagnaigp.it/wp-content/uploads/2023/11/Disciplinare_Pera_dell_Emilia_Romagna_IGP.pdf. (Accessed 28 October 2025).

[16] D. Fornasiero, D. Scaccini, V. Lombardo, G. Galli, A. Pozzebon, Effect of exclusion net timing of deployment and color on *Halyomorpha halys* (Hemiptera: pentatomidae) infestation in pear and apple orchards, Crop Prot. 172 (2023) 106331, https://doi.org/10.1016/j.cropro.2023.106331.

[17] L. Moore, P. Tirello, D. Scaccini, M.D. Toews, C. Duso, A. Pozzebon, Characterizing damage potential of the brown marmorated stink bug in cherry orchards in Italy, Entomol. Gen. 39 (2019) 271–283, https://doi.org/10.1127/entomologia/2019/0799.

[18] V. Cortés, J. Blasco, N. Aleixos, S. Cubero, P. Talens, Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: a review, Trends Food Sci. Technol. 85 (2019) 138–148, https://doi.org/10.1016/j.tifs.2019.01.015.

[19] B. Zhang, D. Dai, J. Huang, J. Zhou, Q. Gui, F. Dai, Influence of physical and biological variability and solution methods in fruit and vegetable quality nondestructive inspection by using imaging and near-infrared spectroscopy techniques: a review, Crit. Rev. Food Sci. Nutr. 58 (2018) 2099–2118, https://doi.org/10.1080/10408398.2017.1300789.

[20] D. Wu, D.-W. Sun, Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: a review — part I: fundamentals, Innov. Food Sci. Emerg. Technol. 19 (2013) 1–14, https://doi.org/10.1016/j.ifset.2013.04.014.

[21] J.M. Amigo, I. Martí, A. Gowen, Chapter 9 - hyperspectral imaging and chemometrics: a perfect combination for the analysis of food structure, composition and quality, in: F. Marini (Ed.), Data Handl. Sci. Technol., Elsevier, 2013, pp. 343–370, https://doi.org/10.1016/B978-0-444-59528-7.00009-0.

[22] A. Gowen, C. Odonnell, P. Cullen, G. Downey, J. Frias, Hyperspectral imaging – an emerging process analytical tool for food quality and safety control, Trends Food Sci. Technol. 18 (2007) 590–598, https://doi.org/10.1016/j.tifs.2007.06.001.

[23] C. Ferrari, G. Foca, R. Calvini, A. Ulrici, Fast exploration and classification of large hyperspectral image datasets for early bruise detection on apples, Chemom. Intell. Lab. Syst. 146 (2015) 108–119, https://doi.org/10.1016/j.chemolab.2015.05.016.

[24] M. Jiang, Y. Li, J. Song, Z. Wang, L. Zhang, L. Song, B. Bai, K. Tu, W. Lan, L. Pan, Study on black spot disease detection and pathogenic process visualization on winter jujubes using hyperspectral imaging system, Foods 12 (2023) 435, https://doi.org/10.3390/foods12030435.

[25] Y. Li, S. You, S. Wu, M. Wang, J. Song, W. Lan, K. Tu, L. Pan, Exploring the limit of detection on early implicit bruised 'Korla' fragrant pears using hyperspectral imaging features and spectral variables, Postharvest Biol. Technol. 208 (2024) 112668, https://doi.org/10.1016/j.postharvbio.2023.112668.

[26] D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, O.L. García-Navarrete, J. Blasco, Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment, Food Bioprocess Technol. 5 (2012) 1121–1142, https://doi.org/10.1007/s11947-011-0725-1.

[27] Y. Lu, Y. Huang, R. Lu, Innovative hyperspectral imaging-based techniques for quality evaluation of fruits and vegetables: a review, Appl. Sci. 7 (2017) 189, https://doi.org/10.3390/app7020189.

[28] N.K. Mahanti, R. Pandiselvam, A. Kothakota, P. Ishwarya S, S.K. Chakraborty, M. Kumar, D. Cozzolino, Emerging non-destructive imaging techniques for fruit damage detection: image processing and analysis, Trends Food Sci. Technol. 120 (2022) 418–438, https://doi.org/10.1016/j.tifs.2021.12.021.

[29] G. Wan, J. He, X. Meng, G. Liu, J. Zhang, F. Ma, Q. Zhang, D. Wu, Hyperspectral imaging technology for nondestructive identification of quality deterioration in fruits and vegetables: a review, Crit. Rev. Food Sci. Nutr. (2025) 1–30, https://doi.org/10.1080/10408398.2025.2487134.

[30] N.-N. Wang, D.-W. Sun, Y.-C. Yang, H. Pu, Z. Zhu, Recent advances in the application of hyperspectral imaging for evaluating fruit quality, Food Anal. Methods 9 (2016) 178–191, https://doi.org/10.1007/s12161-015-0153-3.

[31] J. Wieme, K. Mollazade, I. Malounas, M. Zude-Sasse, M. Zhao, A. Gowen, D. Argyropoulos, S. Fountas, J. Van Beek, Application of hyperspectral imaging systems and artificial intelligence for quality assessment of fruit, vegetables and mushrooms: a review, Biosyst. Eng. 222 (2022) 156–176, https://doi.org/10.1016/j.biosystemseng.2022.07.013.

[32] J. Burger, A. Gowen, Data handling in hyperspectral image analysis, Chemom. Intell. Lab. Syst. 108 (2011) 13–22, https://doi.org/10.1016/j.chemolab.2011.04.001.

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

[34] W.-H. Lee, M.S. Kim, H. Lee, S.R. Delwiche, H. Bae, D.-Y. Kim, B.-K. Cho, Hyperspectral near-infrared imaging for the detection of physical damages of pear, J. Food Eng. 130 (2014) 1–7, https://doi.org/10.1016/j.jfoodeng.2013.12.032.

[35] D. Yu, T. Xu, K. Song, Nondestructive testing of external defects on Nanguo pear, Spectrosc. Lett. 51 (2018) 191–197, https://doi.org/10.1080/00387010.2018.1447968.

[36] Q. Zhang, W. Huang, Q. Wang, J. Wu, J. Li, Detection of pears with moldy core using online full-transmittance spectroscopy combined with supervised classifier comparison and variable optimization, Comput. Electron. Agric. 200 (2022) 107231, https://doi.org/10.1016/j.compag.2022.107231.

[37] S.R. Delwiche, I. Baek, M.S. Kim, Does spatial region of interest (ROI) matter in multispectral and hyperspectral imaging of segmented wheat kernels? Biosyst. Eng. 212 (2021) 106–114, https://doi.org/10.1016/j.biosystemseng.2021.10.003.

[38] A. Gowen, J.-L. Xu, A. Herrero-Langreo, Comparison of spectral selection methods in the development of classification models from visible near infrared hyperspectral imaging data, J. Spectr. Imaging (2019) a4, https://doi.org/10.1255/jsi.2019.a4.

[39] V. Ferrari, R. Calvini, C. Menozzi, E. Costi, D. Giannetti, P. Hoffermans, L. Maistrello, A. Ulrici, NIR hyperspectral imaging to identify damage caused by *Halyomorpha halys* on pears: automated identification of regions of interest related to punctured areas, Spectrochim. Acta. A. Mol. Biomol. Spectrosc. 343 (2025) 126543, https://doi.org/10.1016/j.saa.2025.126543.

[40] R. Calvini, J.M. Amigo, 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. Chim. Acta 967 (2017) 33–41, https://doi.org/10.1016/j.aca.2017.03.011.

[41] M.Á. Martínez-Domingo, E.M. Valero-Benito, J. Hernández-Andrés, Multispectral and hyperspectral imaging, in: A.M. Jiménez-Carvelo, A. Arroyo-Cerezo, L. Cuadros-Rodríguez (Eds.), Non-Invasive Non-destr. Methods Food Integr, Springer Nature Switzerland, Cham, 2024, pp. 175–201, https://doi.org/10.1007/978-3-031-76465-3_9.

[42] R. Calvini, G. Orlandi, G. Foca, A. Ulrici, Development of a classification algorithm for efficient handling of multiple classes in sorting systems based on hyperspectral imaging, J. Spectr. Imaging (2018) a13, https://doi.org/10.1255/jsi.2018.a13.

[43] J.C. Bergh, S.V. Joseph, B.D. Short, M. Nita, T.C. Leskey, Effect of pre-harvest exposures to adult *Halyomorpha halys* (Hemiptera: pentatomidae) on feeding injury to apple cultivars at harvest and during post-harvest cold storage, Crop Prot. 124 (2019) 104872, https://doi.org/10.1016/j.cropro.2019.104872.
[44] H.R. El-Ramady, É. Domokos-Szabolcsy, N.A. Abdalla, H.S. Taha, M. Fári, Postharvest management of fruits and vegetables storage, in: E. Lichtfouse (Ed.), Sustain. Agric. Rev., Springer International Publishing, Cham, 2015, pp. 65–152, https://doi.org/10.1007/978-3-319-09132-7_2.
[45] FAO, Manual for the Preparation and Sale of Fruits and Vegetables: from Fi Eld to Market, Food and Agriculture Organization of the United Nations, Rome, 2004.
[46] P. Gonzalez, J. Pichette, B. Vereecke, B. Masschelein, A. Lambrechts, L. Krasovitski, L. Bikov, An extremely compact and high-speed line-scan hyperspectral imager covering the SWIR range, in: N.K. Dhar, A.K. Dutta (Eds.), Image Sens. Technol. Mater. Devices Syst, Appl. V, SPIE, Orlando, United States, 2018, p. 19, https://doi.org/10.1117/12.2304918.
[47] R. Calvini, G. Foca, A. Ulrici, Data dimensionality reduction and data fusion for fast characterization of green coffee samples using hyperspectral sensors, Anal. Bioanal. Chem. 408 (2016) 7351–7366, https://doi.org/10.1007/s00216-016-9713-7.
[48] V. Ferrari, R. Calvini, C. Menozzi, A. Ulrici, M. Bragolusi, R. Piro, A. Tata, M. Suman, G. Foca, Addressing adulteration challenges of dried oregano leaves by NIR HyperSpectral imaging, Chemom. Intell. Lab. Syst. 249 (2024) 105133, https://doi.org/10.1016/j.chemolab.2024.105133.
[49] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemom. 17 (2003) 166–173, https://doi.org/10.1002/cem.785.
[50] A.L. Pomerantsev, O.Ye Rodionova, Multiclass partial least squares discriminant analysis: taking the right way—A critical tutorial, J. Chemom. 32 (2018) e3030, https://doi.org/10.1002/cem.3030.
[51] R. Calvini, A. Ulrici, J.M. Amigo, Sparse-based modeling of hyperspectral data, in: Data Handl. Sci. Technol., Elsevier, 2016, pp. 613–634, https://doi.org/10.1016/B978-0-444-63638-6.00019-X.
[52] R. Calvini, J.M. Amigo, Coupling randomisation and sparse modelling for the exploratory analysis of large hyperspectral datasets, Chemom. Intell. Lab. Syst. 248 (2024) 105118, https://doi.org/10.1016/j.chemolab.2024.105118.
[53] P. Filzmoser, M. Gschwandtner, V. Todorov, Review of sparse methods in regression and classification with application to chemometrics, J. Chemom. 26 (2012) 42–51, https://doi.org/10.1002/cem.1418.
[54] M.A. Rasmussen, R. Bro, A tutorial on the Lasso approach to sparse modeling, Chemom. Intell. Lab. Syst. 119 (2012) 21–31, https://doi.org/10.1016/j.chemolab.2012.10.003.
[55] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc., Ser. B Stat. Methodol. 58 (1996) 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
[56] D. Ballabio, F. Grisoni, R. Todeschini, Multivariate comparison of classification performance measures, Chemom. Intell. Lab. Syst. 174 (2018) 33–44, https://doi.org/10.1016/j.chemolab.2017.12.004.
[57] D.A. Burns, E.W. Ciurczak (Eds.), Handbook of Near-Infrared Analysis, third ed., CRC Press, 2007 https://doi.org/10.1201/9781420007374.
[58] P. Williams, M. Manley, J. Antoniszyn, Near Infrared Technology: Getting the Best out of Light, AFRICAN SUN MeDIA, 2019.
[59] Jr. Workman Jerry, L. Weyer, Practical Guide to Interpretive Near-Infrared Spectroscopy, first ed., CRC Press, 2007 https://doi.org/10.1201/9781420018318.
[60] S. Gacnik, D. Rusjan, M. Mikulic-Petkovsek, Metabolic response of peach fruit to invasive brown marmorated stink bug (*Halyomorpha halys* Stål.)'s infestation, Int. J. Mol. Sci. 25 (2024) 606, https://doi.org/10.3390/ijms25010606.
[61] I.O. Ozdemir, O. Karakaya, U. Ates, B. Ozturk, M. Uluca, C. Tuncer, Characterization of hazelnut kernel responses to brown marmorated stink bug [*Halyomorpha halys* Stal (Hemiptera: Pentatomidae)] infestations: Changes in bioactive compounds and fatty acid composition, J. Food Compos. Anal. 124 (2023) 105696, https://doi.org/10.1016/j.jfca.2023.105696.
[62] N.C. Weber, J. Razinger, J. Jakopič, V. Schmitzer, M. Hudina, A. Slatnar, R. Veberič, F. Štampar, T. Zamljen, Brown Marmorated stink bug (*Halyomorpha halys* stål.) attack induces a metabolic response in Strawberry (fragaria × ananassa duch.) fruit, Horticulturae 7 (2021) 561, https://doi.org/10.3390/horticulturae7120561.
[63] T. Zamljen, A. Medič, R. Veberič, M. Hudina, F. Štampar, A. Slatnar, Apple fruit (Malus domestica Borkh.) metabolic response to infestation by invasive brown marmorated stink bug (*Halyomorpha halys* Stal.), Horticulturae 7 (2021) 212, https://doi.org/10.3390/horticulturae7080212.
[64] J.-M. Celton, D. Chagné, S.D. Tustin, S. Terakami, C. Nishitani, T. Yamamoto, S. E. Gardiner, Update on comparative genome mapping between Malus and Pyrus, BMC Res. Notes 2 (2009) 182, https://doi.org/10.1186/1756-0500-2-182.
[65] K. Kubitzki, Flowering Plants. Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales, Springer Science & Business Media, 2013.
[66] L. Zapponi, S.G. Chiesa, G. Anfora, L. Chini, L. Gallimbeni, C. Ioriatti, V. Mazzoni, A. Saddi, G. Angeli, Combined effect of timing, position and management on Halyomorpha halys feeding injury assessment on apple cultivars, J. Appl. Entomol. 148–9 (2024) 1027–1036, https://doi.org/10.1111/jen.13319.