Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 343 (2025) 126543

Contents lists available at ScienceDirect

Spectrochimica Acta Part A:
Molecular and Biomolecular Spectroscopy

journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy

NIR hyperspectral imaging to identify damage caused by *Halyomorpha halys* on pears: Automated identification of Regions of Interest related to punctured areas☆

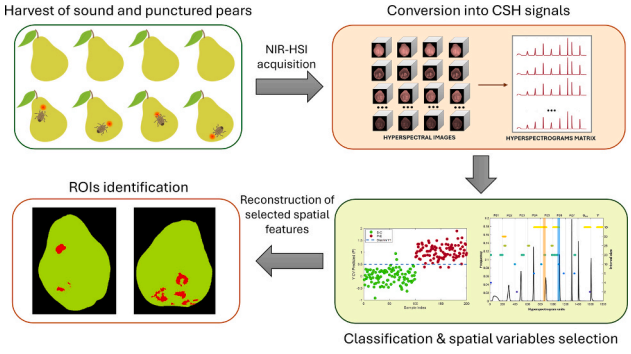
Veronica Ferrari [a], Rosalba Calvini [a,b,*], Camilla Menozzi [a], Elena Costi [a,b], Daniele Giannetti [a], Peter Hoffermans [c], Lara Maistrello [a,b], Alessandro Ulrici [a,b]

[a] Department of Life Sciences, University of Modena and Reggio Emilia, Pad. Besta, Via Amendola, 2, 42122 Reggio Emilia, Italy
[b] Interdepartmental Research Centre BIOGEST-SITEIA, University of Modena and Reggio Emilia, Piazzale Europa, 1, Reggio Emilia 42122, Italy
[c] IMEC OnePlanet, Bronland 10, Wageningen, the Netherlands

HIGHLIGHTS

- NIR-HSI is able to detect under-peel punctures on pears during post-harvest.
- An innovative method for objective and automated ROIs identification is proposed.
- CSH coupled with iPLS-DA selected the spatial features related to punctures.
- ROIs ascribable to punctures were reconstructed on 77.3% images of punctured pears.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Hyperspectral imaging
Data dimensionality reduction
Regions of Interest
Fruit punctures
Post-harvest sorting

ABSTRACT

Halyomorpha halys, commonly known as the Brown Marmorated Stink Bug (BMSB), is an emerging pest in pear orchards determining major economic losses. BMSB feeding on fruits close to harvest ripening cause internal damage invisible to the naked eye, therefore undetectable using RGB image acquisition systems. To face this issue, in the present work Near-Infrared Hyperspectral Imaging (NIR-HSI) is proposed as a non-destructive technique to automatically discard damaged fruits in post-harvest sorting lines.

In this context, the identification of Regions of Interest (ROIs) ascribable to the punctures is a crucial step affecting the outcomes of supervised classification models. Due to irregular shapes and blurred edges between sound and punctured areas, most popular thresholding techniques are not able to automatically detect the ROIs while, on the other hand, manual thresholding is arbitrary and time consuming on large hyperspectral image datasets.

This paper provides an innovative method for the automated ROIs selection based on image data dimensionality reduction (DDR) and image-level classification coupled with spatial feature selection. To this aim, the

☆ This article is part of a special issue entitled: 'IASIM-2024 Special Issue' published in Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy.

* Corresponding author.
E-mail address: rosalba.calvini@unimore.it (R. Calvini).

https://doi.org/10.1016/j.saa.2025.126543
Received 25 March 2025; Received in revised form 19 May 2025; Accepted 5 June 2025
Available online 10 June 2025

hyperspectral images were compressed into Common Space Hyperspectrograms (CSH), signals summarising both spatial and spectral information of the original images. The CSH features highly correlated with the presence of BMSB punctures and more frequently selected by interval Partial Least Squares – Discriminant Analysis (iPLS-DA) models allowed the identification of ROIs of punctured areas. Indeed, the reconstruction of the selected features back into the original image domain led to a successful identification of ROIs ascribable to BMSB punctures in an automated and objective way.

1. Introduction

Halyomorpha halys (Hemiptera, Pentatomidae), also known as the Brown Marmorated Stink Bug (BMSB), is native to East Asia and has become an invasive pest in North America and Europe, posing a significant threat to agriculture. This polyphagous pest attacks a wide range of crops, including fruit, vegetables and ornamentals, leading to significant economic losses [1,2]. In Europe, where BMSB has spread rapidly in almost all countries since 2004, it is causing considerable damage, particularly to fruit crops and hazelnut trees [3]. In northern Italy, BMSB has caused severe damage to fruit crops, with pears being particularly susceptible to this pest [3,4].

BMSB primarily inflicts damage on fruits through its piercing-sucking mouthparts during feeding, causing types of injuries such as deformities, discoloration and internal damage [5–7].

BMSB punctures that occur at a late stage of pear development often result in internal damage that is not externally visible, manifested by the presence of suberified or necrotic areas in the pulp, characterized by a brownish, pithy and corky tissue beneath the skin [7]. Suberification and necrosis significantly reduce the marketability of pears, as they lead to unattractive, poor-quality fruit that is often discarded. Furthermore, internal damage is particularly problematic as it can be extensive yet remaining hidden beneath the fruit skin, making it difficult to detect by visual inspection until the fruit is cut or begins to rot [1].

Near Infrared Hyperspectral Imaging (NIR-HSI) has rapidly emerged as an analytical tool for fruit quality evaluation, allowing to obtain in a fast, cheap and non-destructive manner a detailed insight into the chemical composition and its variation over the sample surface [8–11].

The possibility of coupling both spatial and spectral information makes NIR-HSI particularly suitable for the early detection of fruit defects that are not visible to the naked eye, such as bruises, punctures and various types of damage [12–16]. Furthermore, NIR hyperspectral or multispectral cameras can easily be used for online quality control in post-harvest sorting lines to automatically discard damaged or not compliant fruit.

Given the great potential of this technique for fruit defect detection, in this study we applied NIR-HSI to identify damage caused by BMSB 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. This kind of approach is known as pixel-level classification.

Therefore, pixel-level classification models able to discriminate sound and punctured pear areas must be trained using a library of representative spectra belonging to both classes [17,18]. The selection of such spectra is a crucial aspect for the effectiveness and robustness of the classification models, and it is usually performed by extracting the pixel spectra from Regions Of Interest (ROIs) of the investigated classes [19,20]. Therefore, ROIs identification and labelling are extremely important image elaboration steps affecting the outcomes of supervised classification models. In image analysis, this procedure is often referred to as image annotation and it can be performed using different strategies according to the problem under investigation [21].

Image segmentation based on thresholding procedures is one of the most common approaches to perform ROIs selection, due to its simplicity and efficiency [22,23]. Usually, the grey scale image acquired at the wavelength showing the highest contrast between the area of interest and the remainder pixels is selected, and the corresponding histogram is used to determine a threshold value able to select the pixels of the ROIs.

According to the problem under investigation, a single wavelength may not be sufficient to perform ROIs segmentation, but it is necessary to consider the information brought by different wavelengths at the same time. In this case, it is possible to apply Multivariate Image Analysis (MIA) by calculating a Principal Component Analysis (PCA) model on the investigated image. In this manner, ROIs selection can be performed by thresholding procedures on the score images or by selecting clusters of pixels in the score plots [24,25].

The identification of the proper threshold value is therefore crucial for ROIs selection. The threshold value can be identified either manually by visual inspection of the histogram of grey scale images (i.e., images of single wavelengths or PCA score images) or using automated algorithms [26].

Manual identification of the threshold value can be quite easy when there is a clear separation between ROIs and the remainder parts of the image or when ROIs correspond to specific objects located in the images with a distinctive shape. However, very often ROIs cannot be clearly distinguished and manual thresholding can be a challenging task that may be affected by the operator choice, since different operators can identify different threshold values.

Furthermore, when many images have to be analyzed altogether it is very difficult to define a threshold value appropriate for all the images of the dataset, and manually defining the threshold for each image can be very laborious and time consuming.

Given the drawbacks of manual ROIs selection, it is preferable to use automated procedures [27,28]. Otsu algorithm [29] is the most popular automated thresholding method due to its effectiveness and simplicity. This algorithm identifies a threshold value that separates the pixels of an image in two classes. The threshold value is defined by minimizing the intra-class variance and, at the same time, maximizing inter-class variance.

Munera et al. [30] used an adaptive thresholding based on Otsu method to automatically select ROIs of damaged areas on hyperspectral images of persimmon fruits. This methodology was applied on PC6 score images and then the pixel spectra extracted from the ROIs of damaged areas were used to train supervised classification models.

However, Otsu algorithm provides satisfactory results only when the grey scale image histogram is bimodal and the two groups of pixels are easily separated [31,32]. This condition is rarely met in practical applications of hyperspectral image analysis for fruit damage detection since the distinction between damaged and sound areas is often unclear, in particular when the samples are in an early damage stage.

Improved modifications of the Otsu method or different thresholding algorithms have been applied for automatic segmentation of bruises or other types of injuries in fruits [33–36]. However, they require some assumptions about pixel distributions to be fulfilled or the optimization of several image elaboration steps. Due to these challenges, manual ROIs selection is still a common practice for fruit damage segmentation [14,37].

In this study, the identification of ROIs related to pears areas damaged by BMSB resulted a very challenging task. First of all, it has to be considered that during the experimental part of the study we acquired about 2000 hyperspectral images of sound and punctured pears, also considering different acquisition times from harvest up to five weeks after.

By a preliminary investigation of some representative hyperspectral images using PCA, it was possible to visually identify the pixels related to punctured areas thanks to the observation of the score images. However, there was not a clear distinction between score values of sound and punctured areas. Furthermore, it was not possible to manually find interval(s) of PCA score values applicable to all the images of the dataset, due to the heterogeneity between the different fruits and the variability caused by the varying ripening levels over the acquisition times.

In addition, the shape of punctured areas was irregular and different from fruit to fruit, therefore it was not possible to couple thresholding with object or shape detection algorithms. Moreover, considering the high number of images acquired in this study, it was not feasible to manually identify the ROIs by elaborating each single image.

To face all these challenges, this paper is focused on the development of an innovative method for the automated identification of ROIs based on image data dimensionality reduction and image-level classification coupled with spatial feature selection.

In a previous study developed by some of the authors [16], a large dataset of hyperspectral images of sound and bruised apples was firstly subjected to data dimensionality reduction using the Single Space Hyperspectrograms (SSH) method and then the resulting data matrix was analyzed using interval Partial Least Squares Discriminant Analysis Algorithm (iPLS-DA) [38]. SSH approach consists in reducing the relevant spectral and spatial information of each image into a one-dimensional signal, which is obtained by merging in sequence the frequency distribution curves of quantities derived from a PCA model calculated separately for each image [39]. iPLS-DA applied to the SSH data matrix allowed to perform image-level classification between sound and bruised fruits and to select the hyperspectrogram variables more correlated with the presence of the bruises. The selected variables were then visualized back into the original image domain obtaining the exact localization of bruised areas.

The algorithm for automated ROIs identification developed in this study follows a similar workflow. Firstly, the hyperspectral images of sound and punctured pears were converted into Common Space Hyperspectrograms (CSH) [40], a modification of SSH method. CSH are based on the same principle as SSH, but the signals are obtained by merging in sequence the frequency distribution curves of score values of a global PCA model common to all the images of the dataset used to train the subsequent classification models. Then, iPLS-DA models considering different interval sizes were calculated for image-level classification of sound and punctured fruits. The spatial variables selected more frequently from the iPLS-DA models, corresponding to specific image pixels, were then reconstructed back into the image domain, allowing an automated and objective selection of the pixels related to punctured areas.

2. Materials and methods

2.1. Experimental protocol

The pear samples considered in this study were harvested from an organic orchard located in Carpi (Modena, Italy). The fruits belonged to *Williams* variety and were harvested in summer 2022 and summer 2023.

In both years, 40 tree branches were selected from different plants located in different areas of the orchard. Immediately after fruit set, the selected branches were covered with cylindrical inclusion cages made of a semi-rigid plastic mesh covered with a sleeve of flexible white fabric, each one containing 2 to 4 fruits. The inclusion cages were used to protect the ripening fruit from uncontrolled biotic and abiotic adversities [7,41].

On the same day of commercial pear harvest (August 7th in 2022 and August 4th in 2023), BMSB specimens were manually placed inside half of the inclusion cages. All progenitors of these BMSB specimens had been previously captured in the field from urban parks of the city of Reggio Emilia (Italy) using the tree-beating technique. The bugs were reared in climatic chambers at 26 °C, 60 % relative humidity, L16:D8 photoperiod inside clear mesh cages with organic tomatoes, carrots, green bean pods and raw peanuts as food. A bottle cap with a water-soaked cotton swab was used as water supply. Food and water were replaced twice per week. Feeding of BMSB specimens was interrupted two days before placing them in the inclusion cages.

More in detail, 3 BMSB specimens were placed inside each one of the 20 inclusion cages used to expose the fruits to the bugs. In this manner, about half of the fruits considered were exposed to BMSB feeding punctures, while the remaining fruits were used as control fruits.

In the 2022 season, the bugs were kept in the inclusion cages for two days, from August 7th until August 9th, and then both the exposed and the control fruits were harvested. In particular, 53 control fruits and 40 exposed fruits were harvested in 2022, for a total of 93 pear samples. Unfortunately, only 11 of the 40 exposed fruits were found to have damage ascribable to BMSB punctures. Indeed, the summer of 2022 was characterised by exceptionally high temperatures which reached up to 40 °C at the beginning of August. These extreme climatic conditions negatively affected BMSB specimens' vitality in the field and their propensity to feed.

Based on the outcomes of the summer 2022 harvest, some minor changes were made to the experimental protocol carried out in summer 2023. In this case, the BMSB specimens were kept in the inclusion cages for five days, from August 4th to August 8th. All the fruits were then left on the tree for one week before harvesting, which took place on August 16th. In the 2023 season, a total of 68 fruits were collected: 38 control fruits and 30 BMSB exposed fruits. Considering the exposed pears, 18 out of 30 fruits resulted to have damage ascribable to BMSB punctures.

All the collected fruits presented the typical characteristics of *Williams* variety, and they showed a great size variability, with fruit diameter ranging from 35 mm to 75 mm.

2.2. Hyperspectral image acquisition

In both 2022 and 2023 years, the fruit samples were transported to the laboratory for image acquisition immediately after harvest. During transport, the fruits were placed in boxes covered with egg crate foam to prevent damage.

To monitor the evolution of fruit damage over time, 8 subsequent acquisition times were considered from harvest (T1) until five weeks after (T2-T8).

The schedule followed for hyperspectral image acquisition in 2022 and 2023 is reported in Table 1.

Between the different acquisition sessions, the pears were stored at refrigerated temperatures of 0 – 2 °C to simulate post-harvest storage conditions [42,43].

In order to acquire the whole fruit surface, each pear was divided into four longitudinal sections labelled as A, B, C and D (Fig. 1A). Therefore, one hyperspectral image for each section was acquired for a total of four images for each fruit.

The hyperspectral images were acquired using a line-scanning

Table 1
Schedule for exposure of the fruits to BMSB in the orchard and image acquisition times (T1 – T8); T1 corresponds to the harvesting day.

	2022	2023
BMSB exposure	August 7th – August 9th	August 4th – August 8th
T1	August 9th	August 16th
T2	August 11th	August 23rd
T3	August 16th	August 29th
T4	August 22nd	September 5th
T5	August 29th	September 11th
T6	September 5th	September 14th
T7	September 9th	September 18th
T8	September 12th	September 21st

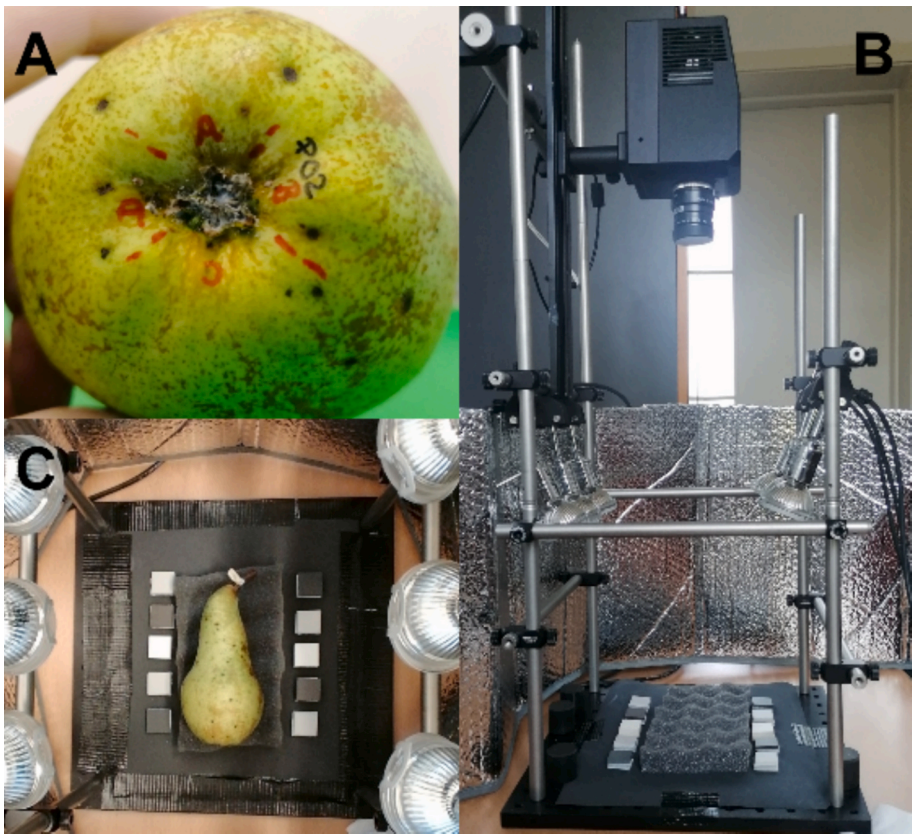


Fig. 1. In A) example of one fruit divided onto four longitudinal sections, in B) and C) image acquisition set-up.

hyperspectral camera (SnapScan SWIR, IMEC One Planet, The Netherlands) working in the 1156 – 1674 nm spectral range with a spectral resolution of 100 spectral channels [44]. The images were acquired considering a lens aperture equal to f/2.8 and an integration time equal to 2 ms. These image acquisition parameters were optimised based on some preliminary tests and kept constant for all the acquisitions performed in 2022 and 2023. Four small halogen lamps with diffusors were utilized for the homogeneous illumination of the samples (Fig. 1B). The acquired images have a spatial resolution of 492 × 820 pixels. Before data acquisition, the camera was calibrated using the acquisition software. This calibration was performed considering the dark current of the device and a white tile with high reflectance.

A black silicon carbide paper was used as image background and the fruits were placed on a black foam rubber holder to prevent them from rolling during image acquisition. The image scene also included 10 ceramic tiles with different grayscale tones and intermediate reflectance values (Fig. 1C). These ceramic tiles were added to correct possible time-depended drifts if necessary. However, the comparison of some representative images acquired at different acquisition times in both 2022 and 2023 years allowed to verify the absence of unwanted variations over time. Therefore, it was not necessary to apply any correction procedure beside standard calibration procedure performed by the acquisition device.

Due to the large number of fruit samples to be acquired at each acquisition time and the need to acquire four images for each pear, corresponding to the four sections, the collected fruits were randomly divided into two groups. The fruits belonging to the first group (Group 1) were acquired at each acquisition time, while those belonging to the second group (Group 2) were acquired only at a specific acquisition time. At each acquisition time, the pears belonging to Group 2 were peeled immediately after hyperspectral image acquisition in order the verify the presence of the damage due to BMSB punctures. Conversely, the pears belonging to Group 1 were peeled only after hyperspectral image acquisition at T8.

At the end of the image acquisition procedure, 952 and 1012 hyperspectral images were acquired in 2022 and 2023, respectively. Therefore, the final dataset of hyperspectral images was composed of 1964 images, corresponding to 1.5 TB.

It was necessary to peel the collected fruits after image acquisition in order to verify the presence of damage due to BMSB punctures. Indeed, this damage affect the fruit pulp, and is not visible at the naked eye on the intact fruits. During this procedure we noticed that not all the fruits exposed to BMSB were actually punctured and, in some cases, signs of damage were not clearly visible and could also be ascribable to mild damage caused by other factors. In addition, some of the control fruits as well as some of the exposed fruits had severe damage due to other biotic agents (e.g., brown spots caused by moulds such as *Stemphylium vesicarium* or *Alternaria spp*., pear scab).

Thanks to the visual inspection of the peeled fruits performed by an expert, it was possible 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.

Superficial discolorations or little black spots, which are typical of Williams variety, were not considered as damage. It was considered as

damage any kind of spot, suberification, necrosis, moulded or deliquescent area present in the fruit pulp, underneath the fruit peel, visibly different from a completely sound area. The number of damaged areas was irrelevant since a fruit section was labelled as damaged if at least one damaged spot was identified.

Table 2 summarises the number of images belonging to the different categories for both 2022 and 2023 harvest years.

Furthermore, based on visual inspection, the damage found in exposed fruits was classified into three categories based on damage type (Table 3):

- Type 1: mild damage of unknown origin;
- 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).

Some examples of the three damage types are reported in Fig. S1 of Supplementary Material.

In some cases, the same fruit section showed several damaged areas pertaining to different damage types. If Type 1 and Type 2 damage were simultaneously found on a fruit section, that section was classified as Type 2. Conversely, if Type 1 or Type 2 damage were found together with a Type 3 damage, the considered section was classified as Type 3.

As reported in Table 3, in 2022 the number of images showing damage related to BMSB punctures (Type 2) was much lower than expected. Indeed, as previously mentioned in Section 2.1, the weather conditions of summer 2022 when the pears were exposed to BMSB were not conducive to bugs vitality and feeding. Conversely, thanks to more favourable weather conditions in summer 2023, it was possible to obtain a higher number of images of fruit sections with punctures due to BMSB.

From here onwards, the images belonging to D-E class with Type 2 damage will be referred to as P-E images (i.e., images of punctured sections of fruits exposed to BMSB).

2.3. Image elaboration

The first step of image elaboration consisted in background removal. To this aim, a PCA model was calculated on each image considering mean center as preprocessing method. All the pixels with negative score values were ascribable to the background composed by the black silicon carbide sheet and the black foam rubber support, and thus they were removed.

An additional step was necessary to remove the pixels ascribable to the plastic label attached on pear pedicel, used to identify the fruit samples. These pixels were characterised by reflectance values higher than 0.5 reflectance units at 1398 nm; therefore, a thresholding procedure was performed considering this wavelength.

Finally, morphological erosion using a disk-shaped structuring element with radius equal to 2 pixels [45] was carried out to remove uninformative pixels belonging to fruit pedicel and calyx [46].

All these steps were performed to each image in an automated manner using routines written *ad hoc* in MATLAB environment (R2020a, The MathWorks, USA) based on Image Processing Toolbox (v. 11.1) and PLS_Toolbox (v. 8.8.1, Eigenvector Research Inc., USA).

Subsequently, *pixel-level* PCA models were calculated on some representative images belonging to P-E class to evaluate the differences between sound and punctured areas. These PCA models were calculated considering linear detrend and mean center as preprocessing methods.

Table 2
Number of images acquired in 2022 and 2023 harvest years of sound sections of control fruits (S-C), damaged sections of control fruits (D-C), sound sections of exposed fruits (S-E) and damaged sections of exposed fruits (D-E). The percentage values are calculated considering for each year the total number of images of control and exposed fruits, respectively.

		2022		2023	
		# of images	%	# of images	%
Control fruits	Sound sections (S-C)	242	53.5 %	340	69.1 %
	Damaged sections (D-C)	210	46.5 %	152	30.9 %
	Total	452	–	492	–
Exposed fruits	Sound sections (S-E)	174	34.8 %	157	30.2 %
	Damaged sections (D-E)	326	65.2 %	363	69.8 %
	Total	500	–	520	–

Table 3
Subdivision of D-E images according to damage type based on visual inspection of the corresponding peeled fruit samples.

		2022		2023	
		# of images	%	# of images	%
Type 1	Mild damage of unknown origin	229	70.2 %	34	9.4 %
Type 2	Damage ascribable to BMSB punctures	54	16.6 %	298	82.1 %
Type 3	Severe damage not ascribable to BMSB	43	13.2 %	31	8.5 %
Total		326	–	363	–

2.4. Method for automated ROIs identification

For the development of the algorithm able to automatically identify the ROIs ascribable to punctured areas, we considered only the images of sound sections of control fruits (S-C) and of punctured sections of exposed fruits (P-E) acquired in 2022 harvest years. This choice was due to technical reasons, since the annotation method was developed between 2022 and 2023 harvesting campaigns. In this manner, the images acquired in 2023 were used in a second moment for a further validation of the method developed in this study for automated identification of punctured areas.

As shown in Table 2 and Table 3, 2022 dataset consists of 242 S-C images and only 54P-E images. Given the large imbalance in the number of available images for the two classes, 159 S-C images out of 242 were randomly selected and used for model development.

Therefore, the final dataset used for the development of the annotation algorithm was composed by 213 images: 159 S-C images and 54P-E images. This dataset of hyperspectral images was divided into training set and test set as follows:

- 150 training images (TR), including 111 S-C images (belonging to 10 fruits) and 39P-E images (belonging to 6 fruits);
- 63 Test images (TS), including 48 S-C images (belonging to 6 fruits) and 15P-E images (belonging to 5 fruits)

The subdivision of the images into TR and TS images was done by keeping into the same set images of the same fruit sample, also when it was acquired at the different acquisition times.

The algorithm developed in this study to perform the automated identification of ROIs can be summarized in the following key steps:

1. Conversion of the images into Common Space Hyperspectrograms (CSH);
2. Calculation of interval Partial Least Squares Discriminant Analysis (iPLS-DA) models on CSH dataset;
3. Image reconstruction of CSH intervals selected by iPLS-DA.

These steps are outlined in Fig. 2, and they will be described in more

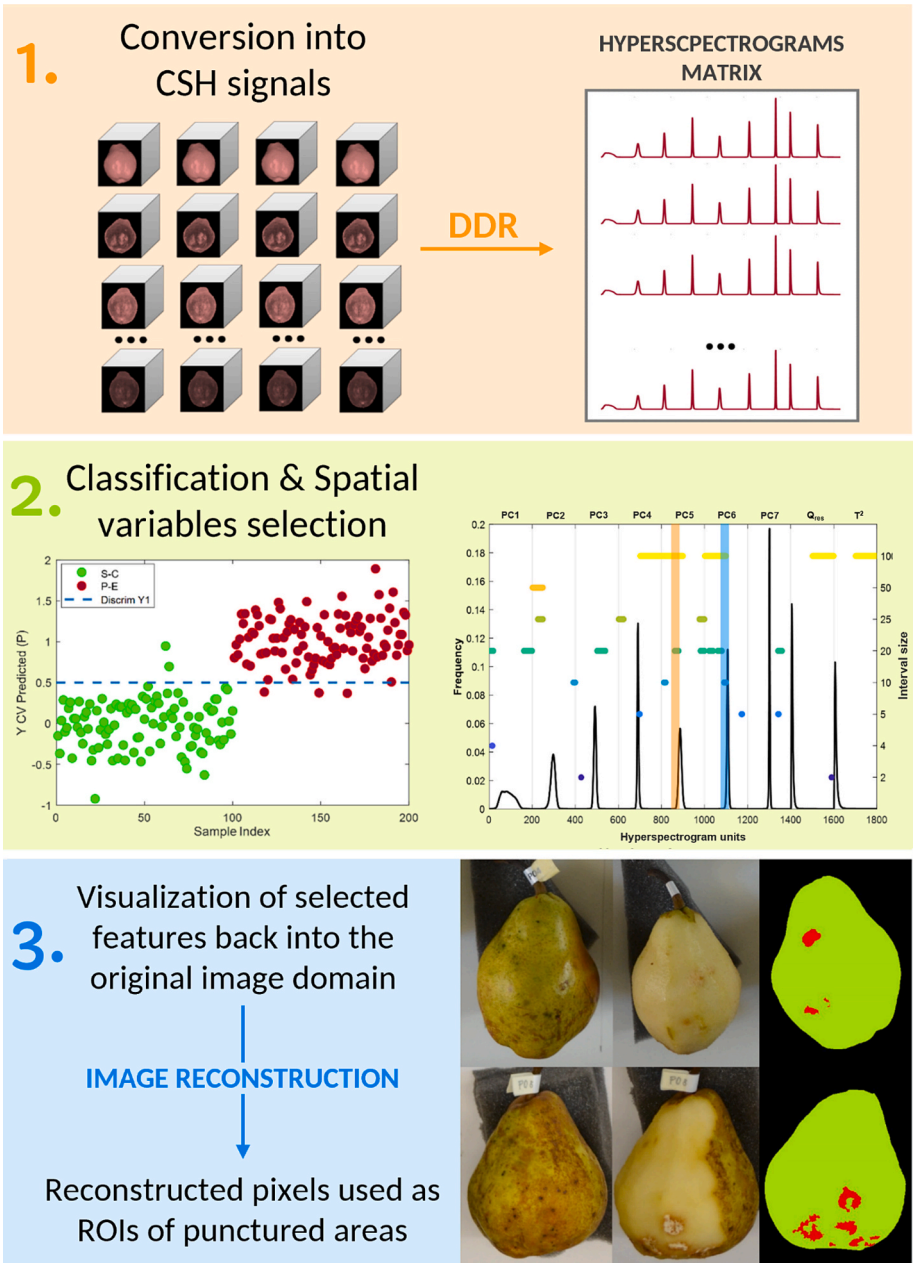


Fig. 2. Schematic representation of the key steps for the automated annotation of ROIs of punctured areas: 1) conversion of the hyperspectral images in CSH signals, 2) application of iPLS-DA for development of image-level classification models (sound vs punctured) and selection of the most relevant variables, 3) visualization of the selected variables back into the original image domain and use of the reconstructed pixels as ROIs ascribable to punctured areas.

detail in the following sections.

2.4.1. Conversion into common Space Hyperspectrograms

Firstly, the hyperspectral images were converted into CSH signals. This operation allows to drastically reduce the dimensionality of the dataset since each image is converted into a features vector, i.e., the hyperspectrogram, which accounts for the relevant spatial and spectral information contained in the corresponding image [39,40,47,48]. CSH signals are generated by combining in sequence the frequency distribution curves of quantities obtained from a PCA model common to all the TR images.

This conversion initially involved unfolding each hyperspectral image into two-dimensional data matrices, where the rows correspond to the pixels retained after background removal and erosion (*see* Section 2.3), and the columns are the spectral variables. Then, the unfolded hypercubes were preprocessed row-wise with linear detrend and centred according to the global average spectrum obtained by averaging all the pixel spectra retained in the TR images. For each image, the variance–covariance matrix was calculated, and the kernel variance–covariance matrix was then obtained by summing the single variance–covariance matrices of all the TR images [49]. The loading vectors of the PCA space common to all the TR images were obtained by applying singular value decomposition (SVD) to the kernel variance–covariance matrix, and 7 PCs were retained.

Subsequently, both TR and TS images were projected onto this common PC space and the corresponding score vectors, Q residuals and Hotelling's T^2 values were obtained. 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 (*see* key step 2 in Fig. 2). To evaluate the effect of the bin size used in the computation of the frequency distribution curves, different CSH datasets were calculated considering bin size values equal to 100 and 200.

Therefore, considering bin size equal to 100 the corresponding TR and TS datasets were composed by 900 variables (= 100 bins × [7 score vectors + Q residuals + T^2]), while for bin size equal to 200 the corresponding TR and TS datasets were composed by 1800 variables (= 200 bins × [7 score vectors + Q residuals + T^2]).

For each bin size, the range of the different frequency distribution curves was separately defined considering the minimum and maximum values of the corresponding quantity across all the TR images. In addition, for each CSH signal the frequency distribution curves were normalised by the number of pixels retained after background removal and erosion.

The CSH signals were calculated using *ad-hoc* routines written in MATLAB environment and based on a graphical user-friendly interface (Hyperspectrograms GUI) freely downloadable from https://www.chimslab.unimore.it/downloads/.

For further details about the procedure used to obtain CSH signals the reader is referred to [40].

2.4.2. Calculation of iPLS-DA models

CSH datasets were analysed by means of iPLS-DA algorithm [38] to discriminate images of sound and punctured fruits, using autoscaling as signal pretreatment. iPLS-DA is a wrapper variable selection method based on the subdivision of the whole signal into a defined number of intervals of equal length. In the *forward* mode used in this study, classification models are iteratively calculated by adding intervals until the minimum classification error in cross-validation is reached [50]. According to the number of bins used to obtain the frequency distribution curves included in CSH signals, different interval sizes for iPLS-DA were tested. For the CSH dataset obtained considering 100 bins, the iPLS-DA models were calculated with interval sizes equal to 50, 25, 20, 10, 5, 4 and 2, while the iPLS-DA models developed for the CSH dataset with 200 bins considered interval sizes equal to 100, 50, 25, 20, 10, 5, 4 and 2.

A Leave-One-Fruit-Out cross-validation scheme was adopted: in each cross-validation iteration all the signals obtained from the images of the same fruit were left out.

The performances of the iPLS-DA models were evaluated in terms of sensitivity (SENS), specificity (SPEC) and efficiency (EFF) in cross-validation and in prediction [51].

However, it has to be highlighted that the iPLS-DA models were calculated with the main aim of selecting in an automated and objective manner intervals of hyperspectrogram variables that are mostly related to the presence of punctured areas. Therefore, the main aim of this part of the study was not to obtain classification models with satisfactory performances, but to check their ability to identify the ROIs ascribable to punctures. This latter task was performed using the image reconstruction procedure explained in the following section.

2.4.3. Image reconstruction of selected variables

Since CSH are obtained by merging in sequence the frequency distribution curves of score, Q residuals and T^2 vectors, performing variable selection on this kind of signals means selecting groups of pixels sharing similar features, which in turn are related to the problem of interest (i.e., the presence of punctures in this study).

For example, suppose that, after variable selection using iPLS-DA, the model selects one interval of 10 variables falling in the CSH region of the frequency distribution curve of PC3 and that the selected variables correspond to PC3 score values ranging from 0.35 to 0.45. Image reconstruction allows to visualise back into the original image domain the selected CSH variables by visualising only the pixels falling in the intervals of interest. Considering the previous example, we are interested in visualising back into the original image domain only the pixels with PC3 score values in the range 0.35 – 0.45. In our case these pixels are ideally ascribable to the punctures and constitute the ROIs.

Further details about the image reconstruction procedure are reported in [39] and [16].

In addition to comparing the ability to identify ROIs by individual iPLS-DA models, also the overall frequency of selection of the CSH regions resulting by the different interval sizes was taken into account. In fact, since the position and width of the useful CSH features are not known in advance, it is advisable to consider different interval size values and then to focus on the regions that are the most frequently selected ones [52]. Therefore, the reconstruction of the CSH regions most frequently selected was also performed by analysing the two CSH datasets (100 bins and 200 bins) separately from each other. In this way it was possible to highlight the most useful CSH regions regardless of the specific iPLS-DA interval size, and to use them to improve ROIs selection.

3. Results and discussion

3.1. Pixel-level exploratory analysis

A preliminary exploratory analysis with PCA was performed on some representative images of P-E fruits harvested in 2022 to assess the spectral differences between pixels ascribable to punctured and sound areas. For this investigation, the segmented hyperspectral images were preprocessed row-wise considering different methods (i.e. SNV, derivatives, detrend) combined with mean centering. The linear-detrend row preprocessing was the one allowing a better separation between pixels related to punctured and sound areas.

According to the outcomes of the PCA models calculated on single P-E images, a clear separation between sound areas and punctures was hard to achieve. First of all, the information related to the presence of actual damage related to BMSB activity represented less than 0.10 % of the explained variance and, secondly, in some images this kind of information was retrieved by different PCs or diverse intervals of PC score values.

As an example, Fig. 3 reports the PCA results obtained from two images of different punctured fruits acquired at T6 and T7 acquisition times, respectively. For both images, PC5 is the component allowing to better locate the punctures thanks to the visualisation of the corresponding PC5 score images and their comparison with the RGB images of peeled fruits. However, observing the PC1 *vs* PC5 score plots and the PC5 histograms, it is clear that pixels of punctured areas do not form a well-defined cluster distinct from the pixels of sound areas.

Using the brushing approach [24], for both images it was possible to manually identify PC5 score values mainly characterising punctured areas thanks to the possibility of interactively visualising the corresponding pixels (highlighted in magenta colour in Fig. 3B–D and G–I). For the image acquired at T6, the pixels of punctured areas fall into the −0.012 – −0.048 PC5 score range (Fig. 3B–D) while for the image acquired at T7 pixels of punctured areas have PC5 score values from −0.018 – −0.040 (Fig. 3G–I). Even if these two intervals partially overlap, it is almost impossible to identify threshold values suitable for both images and the absence of a clear cluster of pixels of the punctures makes it difficult to apply traditional automated thresholding methods like the Otsu algorithm. Moreover, the areas ascribable to punctures are characterized by strongly irregular shapes not particularly dense in pixels, determining blurred edges between sound and punctured areas (Fig. 3 B and G), thus not allowing a clear separation using morphological operators.

Another aspect that further complicated the selection of punctured areas was that in some images, PCs other than PC5 were also involved in punctures detection, making the thresholding procedure even more difficult since multiple PCs had to be considered at the same time.

All these aspects made the correct identification of ROIs ascribable to the punctures a very challenging task. Basically, it would have been necessary to perform PCA at the pixel-level on each image of P-E fruits to manually identify the ROIs, but considering the high number of images acquired, this operation was not feasible. In addition, the threshold selection on relevant score values and related pixels ascribable to ROIs would have been subjective and strictly dependent on the operator.

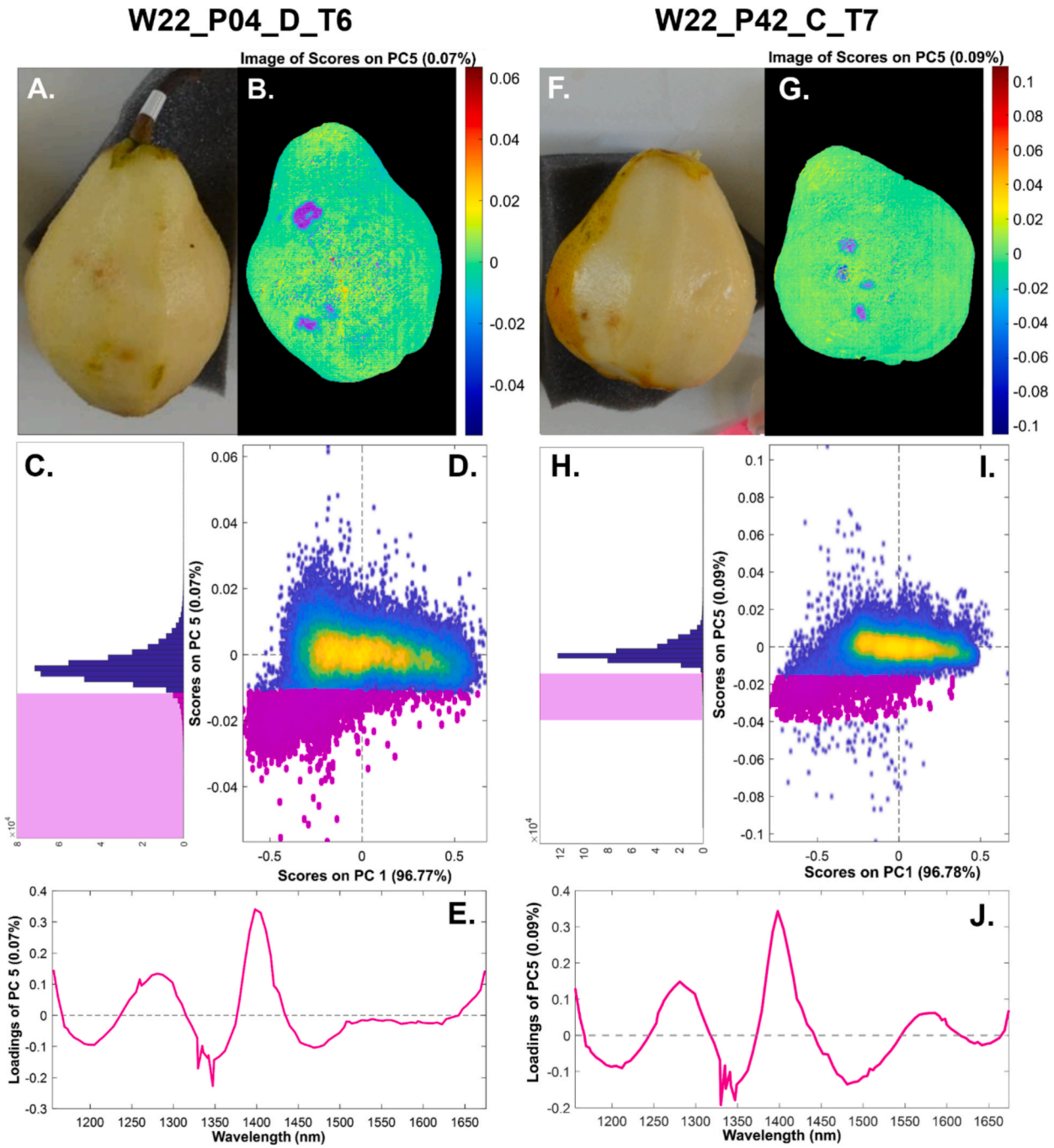


Fig. 3. PCA results of different P-E fruits images acquired at T6 and at T7. In (A, F) the RGB of the peeled fruit used as reference, in (B, G) PC5 score image, in (D, I) PC1-PC5 score plot, in (C, H) histogram of PC5 scores, in (E, J) loading vector of PC5. In (B-D, G-I) the pixels highlighted in magenta correspond to those selected as ROIs of punctured areas using the brushing tool.

Therefore, to perform a selection of ROIs in an objective and automatized way, it was necessary to adopt image-level analysis strategies (*see* Section 2.4).

Interestingly, the spectral information related to the presence of punctures reported by the loadings of PC5 in Fig. 3E and J seemed coherent on the majority of investigated P-E images, for different types of BMSB related damage (i.e. suberifications and necrosis of fruit pulp, which are shown in Fig. 3A and F respectively). The more relevant spectral regions in the loadings of PC5 fall into the 1100–1250 nm spectral range (C – H second overtone), in the 1370–1420 nm spectral range (C – H_2 combination band, aromatic C – H combination band and O – H stretch first overtone), in the 1320–1370 nm (C – H_3 combination band) and in the 1430–1500 nm region (C=O stretch third overtone and N-H stretch first overtone) [53,54]. These outcomes suggest that the spectral differences between damaged and sound areas reflect the different chemical composition of the tissues. Indeed, different studies reported a change of sugars and polyphenols content in fruits as a metabolic response to BMSB punctures [55–57]. Interestingly, Zamljen et al., 2021 [57] demonstrated 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. Nonetheless, since both apple (*Malus* spp.) and pear (*Pyrus* spp.) fruits belong to the *Maloideae* sub-family, they may

exhibit a similar metabolic behaviour toward BMSB punctures [58,59]. It has to be considered that different damaging agents or stresses (e.g., moulds or other diseases) may induce different metabolic responses in fruits, determining different chemical modifications of the fruit pulp. However, these evaluations are out of the aim of current manuscript as this study is focused on identifying damage caused by BMSB punctures.

3.2. Image-level classification and variable selection by iPLS-DA

Starting from CSH datasets calculated considering 100 and 200 bins, both PLS-DA and iPLS-DA classification models were calculated to discriminate between P-E and S-C hyperspectrograms. Considering iPLS-DA, different interval sizes were tested (*see* Section 2.4.2). The classification results of all the calculated models are summarized as heatmap of cross-validation efficiency (EFF CV) values in Fig. S2 of Supplementary Material. Unfortunately, PLS-DA led to overall poor classification results, with EFF CV values equal to 0.381 and 0.417 for the CSH datasets calculated with 100 and 200 bins, respectively.

Moving to iPLS-DA results, it is possible to observe an improvement of the classification results of some models with respect to PLS-DA. More in detail, the iPLS-DA model calculated on the CSH 200 bins dataset with interval size of 20 variables led to the highest EFF CV value, corresponding to 0.766. Acceptable results were obtained also for interval sizes of 50 and 100 variables for CSH 200 bins dataset, and interval sizes of 4, 5 and 10 variables for CSH 100 bins dataset. For each CSH dataset, the three best performing models were selected based on EFF CV values and further investigated by calculating their prediction performances using the corresponding CSH test set.

The cross-validation and test set prediction results of the selected models are reported in Table 4, in terms of SENS, SPEC and EFF values referred to the P-E class.

It is possible to observe that these models generally have high SPEC but poor SENS values for the P-E class, suggesting that they do not correctly recognise the punctured areas of the fruits. Moving to the prediction results, there is a general slight improvement in the model performances, in particular considering the SENS values. The only exception is the model calculated from the CSH 200 bins dataset and considering an interval size of 50 variables, which led to very poor classification results.

Unfortunately, not completely satisfactory results were obtained from the iPLS-DA classification models (Table 4). This outcome can be motivated also considering notable imbalance of samples between the two classes. As shown in Table 3, the sections of fruits exposed to BMSB that actually exhibited punctures (P-E) were significantly lower than the sections of sound control fruits (S-C). Moreover, the overlapping of the two classes and the rather low variance explained by the actual presence of damage may concur to these outcomes (Fig. 3). However, the main focus of this part of the study was not to obtain satisfactory image-level classification results, but to develop an objective and automated method to identify the ROIs ascribable to BMSB damage.

3.3. Rois visualisation based on iPLS-DA selected variables

The hyperspectrogram variables selected by iPLS-DA correspond to intervals of score values that are likely to be related to the P-E class, which in turn are more likely to correspond to the pixels of punctured areas. Indeed, the selected variables can be visualised back into the original image domain using the image reconstruction procedure illustrated in Section 2.4.3.

Based on these considerations, for each one of the best performing iPLS-DA models reported in Table 4, image reconstruction of P-E images was carried out considering the selected hyperspectrograms variables.

The image reconstruction results are shown in Table 4, reporting the number of P-E images for which the ROIs of punctured regions were reconstructed considering the selected intervals of a specific frequency distribution curve composing the CSH signals, and the overall number of reconstructed images.

Considering the CSH 100 bins dataset, the interval size 5 and interval size 10 models allowed to identify the highest number of ROIs for 26 P-E images out of 54; the selected intervals fall in the PC1 and PC5 frequency distribution curves of the CSH signals. These are the best performing models in terms of number of reconstructed images, although they are not the models leading to the best image-level classification results.

While being the best in terms of EFF CV values obtained globally, the iPLS-DA model calculated on the CSH 200 bins dataset considering interval size of 20 variables led to the visualization of ROIs on only 15 images out of 54. Also in this case, the relevant intervals selected fall into the PC1 and PC5 frequency distribution curves of CSH signals.

The iPLS-DA model calculated considering an interval size of 50 intervals on CSH 200 bins dataset led to acceptable results in cross-validation but very poor in prediction. Interestingly, this model led to the selection of only one interval falling in the PC2 scores frequency distribution curves which didn't allow to visualize any ROI.

Lastly, the iPLS-DA model with interval size of 100 variables on CSH 200 bins dataset allowed the visualization of ROIs on 10P-E images out of 54. Although not the best in cross-validation, this model led to the best results in prediction and, more importantly, allowed the ROIs visualization considering an interval falling in the frequency distribution curve of PC6. This hyperspectrogram region has been frequently selected also for other models, but did not result in ROIs visualization.

In order to improve the results in terms of number of P-E images with ROIs of punctures automatically identified, it is crucial to gather all the relevant spatial features (i.e., intervals selected from CSH signals) related to the presence of damage. Therefore, considering the 100 bins and 200 bins CSH datasets separately, the selected intervals were also evaluated based on their frequency of selection by the different iPLS-DA models with varying intervals sizes.

This operation was done based on the assumption that intervals of

Table 4
Classification performances of the three best iPLS-DA models calculated for each CSH dataset (SENS and SPEC values are referred to the P-E class) and corresponding number of P-E images for which the ROIs of punctured regions were successfully reconstructed. Only the selected PCs leading to ROIs identification are reported. In the columns related to the number of images with identified ROIs, "0" indicates that the interval was selected but no ROIs were reconstructed while "-" indicates no interval selection.

			Classification results						Number of images with identified ROIs			
			CV			PRED			PC selected			Tot.
CSH	Interval size	LVs	SENS	SPEC	EFF	SENS	SPEC	EFF	PC1	PC5	PC6	
100 bins	4	6	0.513	0.874	**0.669**	0.600	0.833	**0.707**	7	12	–	**15**
	5	6	0.513	0.863	**0.673**	0.733	0.792	**0.762**	8	24	0	**26**
	10	5	0.538	0.820	**0.664**	0.600	0.813	**0.698**	8	24	0	**26**
200 bins	20	6	0.692	0.847	**0.766**	0.800	0.604	**0.695**	3	12	0	**15**
	50	3	0.462	0.964	**0.667**	0.133	1.00	**0.365**	–	–	–	**0**
	100	5	0.539	0.793	**0.653**	0.867	0.854	**0.860**	–	0	10	**10**

variables selected more frequently are more likely related to the presence of BMSB damage, regardless of the performance of the single models.

Fig. 4 displays the intervals selected by iPLS-DA models calculated on the 100 bins and 200 bins CSH datasets: the frequencies of selection are shown in Fig. 4A and B, while Fig. 4C and D report the variables selected from each iPLS-DA model.

For each CSH dataset we considered the variables selected most frequently by the different iPLS-DA models, and for each interval we visualised the corresponding reconstructed images to verify the correct identification of ROIs related to punctures. The results are reported in Table 5 as number of P-E images acquired in 2022 correctly reconstructed.

Considering CSH 100 bins dataset, the most frequently selected intervals correspond to variables selected three times and they fall in CSH regions of the frequency distribution curves of PC1 and PC5. The reconstruction of these intervals led to the correct visualisation of the ROIs of punctured regions in 26 images out of the 54P-E images acquired in 2022, corresponding to 6 fruits out of 11 for which it was possible to annotate the punctures. For the majority of the reconstructed images, the interval selected on PC5 is the one providing the visualisation of the ROIs, however for some images also the interval selected on PC1 allowed to reconstruct the pixels of punctured areas. This finding is coherent with the outcomes of the single iPLS-DA models reported in Table 4. Variables selected at least twice were also evaluated, but did not result in additional ROIs visualisation.

Table 5
Number of P-E images with reconstructed ROIs obtained considering the most frequently selected intervals of score values for the CSH 100 bins and CSH 200 bins datasets. The number of P-E images is subdivided based on the corresponding PC. The total number of P-E images and P-E fruits with identified ROIs is also reported.

CSH	Frequency	PC1	PC5	PC6	# P-E Images /Tot. 54	# P-E fruits /Tot. 11
200 *bins*	2	0	21	11	**32**	**8**
100 *bins*	3	8	24	–	**26**	**6**

Considering the results obtained for CSH 200 bins dataset, image reconstruction was performed with the variables selected twice by the different iPLS-DA models and corresponding to CSH regions falling in the PC1, PC2, PC5, PC6, PC7, and Q residuals frequency distribution curves. Among these intervals, only those falling in the PC5 and PC6 frequency distribution curves (highlighted in Fig. 4D) led to the correct visualisation of the ROIs of punctured areas in 32 P-E images out of 54 P-E images of 2022 (Table 5), corresponding to 8 fruits out of 11. These interval variables correspond to PC5 score values ranging from −0,025 to −0,016 and PC6 score values ranging from −0,020 e −0,015.

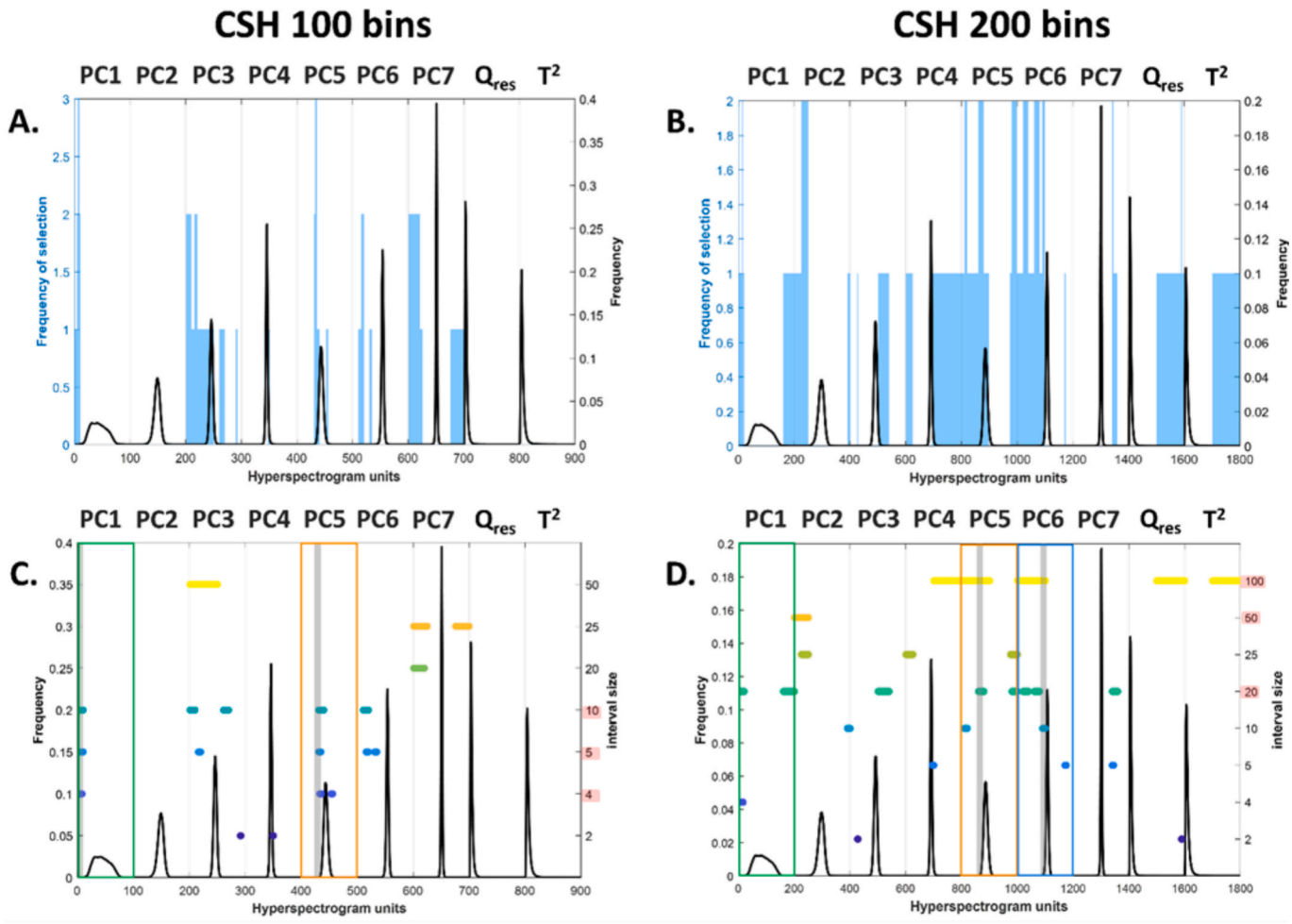


Fig. 4. Frequency of selection of the CSH variables considering the different iPLS-DA models calculated for CSH 100 bins (A) and CSH 200 bins (B) datasets and CSH variables selected by the iPLS-DA models calculated with different interval sizes (C and D). In C) and D) the highlighted interval size values correspond to those leading to the correct visualisation of ROIs of punctured regions using the image reconstruction procedure. The framed frequency distribution curves correspond to the PCs that led to at least one ROI reconstruction (see Table 4).

The selection based on the variables selected most frequently from the different iPLS-DA models on CSH 200 bins dataset allowed to outperform the image reconstruction results obtained by the individual iPLS-DA models. It is likely that a higher number of bins for the calculation of the frequency distribution curves composing the hyperspectrograms eased the separation between spatial features related to pixels of punctured areas and the ones related to sound pixels, thus allowing a better ROIs visualization for the CSH 200 bins dataset with respect to the 100 bins dataset.

Fig. 5 shows some examples of image reconstruction based on the PC5 and PC6 most frequently selected intervals considering the CSH 200 bins dataset. The image reconstruction and ROIs visualisation are shown not only for P-E images, but also for a S-C image, a D-E "Type 3" image (i.e., damaged-exposed with severe damage not ascribable to BMSB), and a D-C image. Furthermore, the examples of ROIs reconstruction are provided with the corresponding RGB images of unpeeled and peeled fruits as reference.

More in detail, Fig. 5 A reports the image reconstruction results obtained for a S-C image (i.e., sound section of a control fruit), where only few sparse pixels falling in the PC6 selected interval were reconstructed back into the original image domain. However, these pixels can be easily filtered out by applying morphological operators to the reconstructed image. In particular, the combination of *erosion* (disk-shaped structuring element with radius of 2 pixels) and *filling holes* morphological operators allowed to completely remove these pixels, as shown in the final ROIs visualization.

On the other hand, the spatial features selected on CSH signals allowed to reconstruct image regions corresponding to damaged areas of P-E images (Fig. 5B–C). Generally, as shown in orange pixels in Fig. 5B, the reconstruction of the selected PC5 interval corresponds to the pixels belonging to punctures, specifically ascribable to the edges between sound and punctured areas. In some other cases, the selected interval on PC6 scores brought relevant information about the presence of punctured areas, and the corresponding reconstructed pixels usually fell inside the edges of the punctured areas (blue pixels in Fig. 5C).

Also in the case of P-E images, after image reconstruction of the selected spatial features of CSH signals, the same morphological operators described above were applied to further optimize ROIs selection. The ROI visualization is reported in red colour in the lowest row of Fig. 5B and C. Comparing the final ROI annotation of punctured areas with the corresponding RGB image of the peeled fruit section, it is possible to verify the correct identification of punctured regions. Furthermore, considering the RGB images of the unpeeled fruit sections it is possible to observe that the damage occurs in the pulp under the peel and they are not visible at the naked eye.

The procedure developed in this study allowed the correct reconstruction of ROIs of both suberified and necrotic areas, which are the typical types of damage caused by BMSB punctures. However, suberification and necrosis of the fruit pulp may also be due to other causes as they are a common response of plants and corresponding fruits to stress. For this reason, we tested the ability of the proposed approach to visualize this damage when it is not caused by BMSB punctures. To this aim, Fig. 5D shows the correct reconstruction of a suberification damage found in a section of a control fruit (D-C), therefore not due to BMSB. Suberified tissues have the same chemical composition regardless of the agents that caused them to emerge, and fruits with suberified areas are considered of low quality. Therefore, we verified that the developed approach is capable of reconstructing damage with the same chemical composition, although not directly caused by BMSB.

Finally, we also tested the proposed approach in identifying damage different from suberification and necrosis, such as mould. Fig. 5E shows one section of a pear exposed to BMSB (D-E) which exhibits both BMSB-induced suberification and a moulded area. In this case, the identified ROI is only related to the suberified area, suggesting that the selected hyperspectrogram intervals are specific only for damage similar to that caused by BMSB.

3.4. Validation of the ROIs selection method

As previously described in Section 2.4, the automated ROIs visualization method was developed based on the images acquired in 2022. Specifically, the CSH intervals falling in the PC5 and PC6 frequency

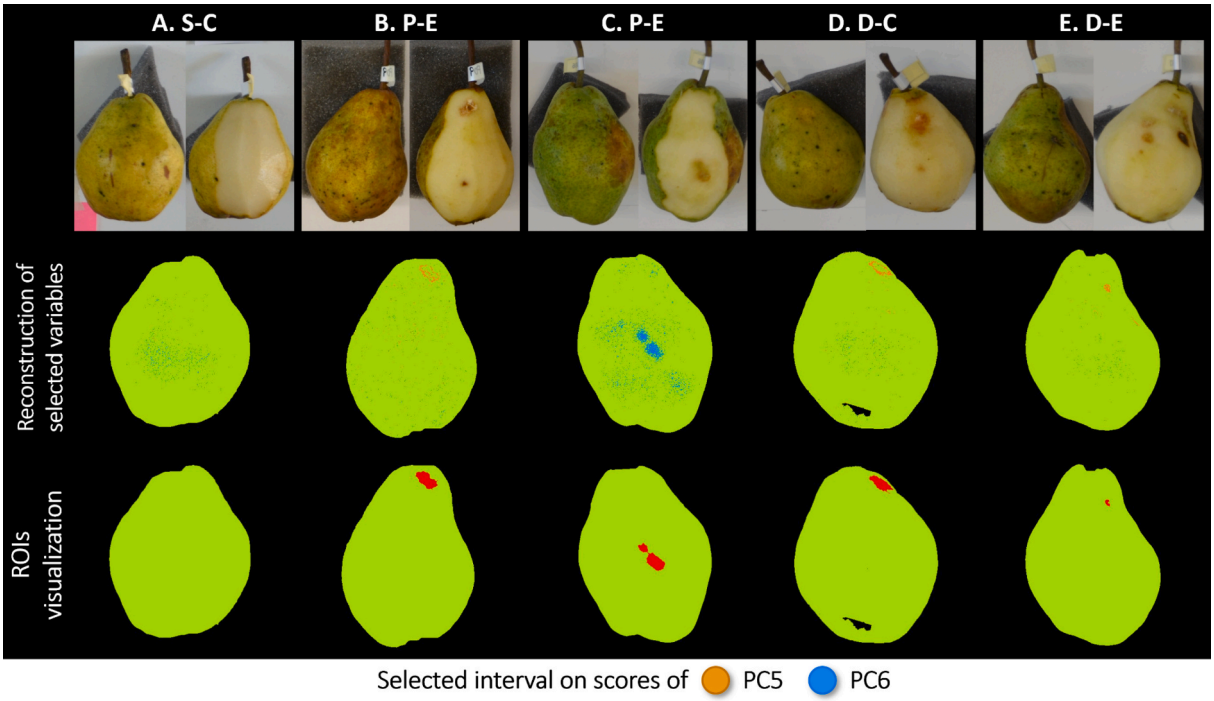


Fig. 5. Reconstruction in the original image domain of PC5 (orange) and PC6 (blue) selected intervals ascribable to punctures for S-C (Sound-Control), P-E (Punctured-Exposed), D-C (Damaged-Control) and D-E (Damaged-Exposed) pears. The final ROIs annotation obtained after erosion and filling holes morphological operators is shown in red colour. Reconstructed images are reported together with the corresponding RGB images of the fruits before and after peeling.

distribution curves used for image reconstruction were selected considering the best results in terms of correctly reconstructed P-E images.

Based on the developed approach, we correctly annotated the ROIs of punctured areas in 32 out of 54 P-E images acquired in 2022, corresponding to 59.2 % (Table 6). Since part of the collected fruits were acquired at different acquisition times, we also evaluated the number of fruits for which we had the correct visualization of damaged areas. Considering 2022 samples, we annotated images corresponding to 8 damaged by BMSB exposed fruits out of 11 fruits (72.7 %).

The method for automated identification of ROIs of punctured areas was further validated considering the P-E hyperspectral images acquired in 2023. The corresponding CSH 200 bins dataset was calculated by projecting the 2023 images on the common PC space calculated for 2022 images (*see* Section 2.4.1); then, the selected intervals on PC5 and PC6 frequency distribution curves were visualised back into the original image domain. As shown in Table 6, satisfactory results were obtained in this validation step, with 240 out of 298 of 2023 P-E images (80.5 %) for which it was possible to correctly reconstruct the ROIs of damaged areas. The outcomes out this validation step are even more satisfactory considering that we were able to correctly visualise the ROIs of punctures in at least one image of all the punctured fruits collected in 2023.

The different performances obtained in ROIs reconstruction for the P-E images acquired in 2022 and 2023 are probably due to the different conditions occurred in the two years. Indeed, climatic conditions in 2022 were not favourable for BMSB feeding activity, resulting in the collection of a limited number of punctured fruits. Conversely, climatic conditions in summer 2023 were more favourable for BMSB vitality, leading to a greater intensity and incidence of damage compared to summer 2022, thus allowing a greater number of P-E fruits to be collected.

The ability of the methodology towards the identification of punctures during post-harvest storage was evaluated on P-E pears acquired at each acquisition time (*see* Section 2.2).

Fig. 6 reports an example of the reconstructed pixels and the final ROIs annotation for P-E pears harvested in 2022 and 2023. For instance, the damaged areas reconstructed by the selected PC5 score interval on P-E images of the pear harvested in 2023 are consistent. The identification of ROIs ascribable to the punctures was feasible at any time, from harvest to six weeks later in post-harvest storage conditions. Nonetheless, the different rotation of the fruits during the subsequent HSI acquisition times had a major impact on the visualization of the ROIs, especially for the image acquired at T2 of P-E harvested in 2022, where the punctured area was inadvertently not imaged.

For both P-E fruits shown in Fig. 6, the reconstruction of the selected PC5 interval in orange corresponds to the pixels belonging to the punctured area, specifically related to the edges between sound and punctured regions. As already observed, PC5 seems retaining the predominant information able to separate the sound areas from the punctured ones, thus allowing the reconstruction of a greater number of P-E images with ROIs correctly identified. In addition, the wavelengths characterized by the highest absolute values of loadings of PC5 calculated on the 200 bins CSH dataset (Fig. S3 of Supplementary Material) are coherent with the higher absolute values of loadings of PC5 obtained with a PCA model calculated on an individual image (Fig. 3E). As described in Section 3.1, the more relevant spectral regions seem to confirm again a possible variation in sugars and polyphenols content in punctured tissues [55–57].

Table S1 of Supplementary Material reports the outcomes for reconstructed images belonging to P-E fruits, subdivided by acquisition time (T1-T8) and harvest year. Overall, the percentage of P-E images harvested in 2022 with ROIs reconstructed is lower than the ones harvested in 2023: the worst performances for 2022 samples are due to the fewer P-E images collected for each acquisition time and to the three P-E fruits for which ROIs were not annotated. In a few cases, the areas ascribable to the punctures are partially or not annotated due to the fragmented reconstruction of ROIs with lower pixels density, which are subsequently discarded as noise.

Table 6
Results obtained from the ROIs visualisation method for P-E images acquired in 2022 and 2023.

	Harvest year	
	2022	**2023**
Reconstructed P-E images	32	240
Total P-E images	54	298
% Reconstructed P-E images	**59.2 %**	**80.5 %**
Reconstructed P-E fruits	8	18
Total P-E fruits	11	18
% Reconstructed P-E fruits	**72.7 %**	**100 %**

4. Conclusions

In the present work, an innovative approach is proposed for the automated and objective identification of ROIs related to punctured areas on NIR hyperspectral images of pears.

In the investigated case study, the identification of one or more threshold values to select the pixels of punctured regions was a challenging task due to the complexity of the dataset.

Indeed, a preliminary pixel-level analysis performed by PCA on individual images highlighted the absence of a clear separation between pixels of punctured regions and pixels of sound areas, therefore not allowing the identification of a unique threshold on PC scores common to all the images. On the other hand, performing manual selection of ROIs by means of the evaluation of PCA models calculated on each single image was not feasible given the high number of images and its arbitrariness.

To overcome these limits, we developed an innovative method for the objective and automated identification of ROIs based on the calculation of Common Space Hyperspectrograms (CSH) and image-level classification coupled with spatial features selection using iPLS-DA. The dimensionality reduction of the image data enabled by CSH was fundamental to handle the large size of the investigated dataset, allowing to compress the relevant information into a global PCA space common to the training set images.

Despite not leading to satisfactory classification results at the image-level, the spatial variables selected more frequently by iPLS-DA models led to the selection of spatial features more correlated to the presence of punctures. Then, these spatial features were visualized back into the original domain, and this image reconstruction procedure, coupled with morphological operators, allowed to correctly annotate the ROIs of punctured areas in the majority of P-E images. Puncture-like damage not related to BMSB activity (e.g., suberifications) was successfully annotated while, vice versa, damage not attributable to BMSB (e.g., mould) was not reconstructed.

Even if the image compression performed by CSH approach can be seen as a loss of pixel-related resolution of the images, the outcomes of this study demonstrate that this approach to data reduction preserves the spatial information, allowing to effectively identify spatially resolved features like damaged spots.

Furthermore, it has to be highlighted that the proposed approach is not to be intended as a classification method, but it was developed only as a strategy to effectively identify ROIs of punctured areas in the hyperspectral images of the collected pears. Starting from the ROIs obtained following this procedure, the second part of the study will focus on the development of pixel-level classification models. Indeed, the selected ROIs will be used to create a dataset of representative spectra belonging to both punctured and sound areas. In turn, the so obtained dataset will be used to develop supervised pixel-level classification models for predictive purposes in post-harvest sorting systems.

More in general, the proposed annotation algorithm can be used in

Fig. 6. Reconstruction in the original image domain of score values of PC5 (orange) and PC6 (blue) ascribable to punctures for P-E (Punctured-Exposed) pears harvested in 2022 and 2023 over time. Along with the RGB references of the fruits before and after peeling, the final reconstructions are shown in red.

all those situations were many images have to be analysed altogether and the identification of common threshold values for ROIs selection is very difficult due the complexity of the considered problem.

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 and Dr. Elena Costi 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 their research (CUP: E95F21002330001).

The authors wish to express their gratitude to Dr. Niccolò Patelli (Applied Entomology Lab, UNIMORE) for the support on-field and Enrico Giovanella for the valuable technical support during the image acquisition and investigation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2025.126543.

Data availability

Data will be made available on request.

References

[1] 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) A1–A13, https://doi.org/10.1603/IPM14002.

[2] 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/CITE/REFWORKS.

[3] L. Maistrello Case Study 2: Halyomorpha halys (Stål) . Europe, in: A.F. Bueno, A.R. Panizzi, Stink Bugs (Hemiptera: Pentatomidae) 2024 Springer Research and Manageme 271 359 10.1007/978-3-031-69742-5_15.

[4] L. Maistrello, G. Vaccari, S. Caruso, E. Costi, S. Bortolini, L. Macavei, G. Foca, A. Ulrici, P.P. Bortolotti, R. Nannini, L. Casoli, M. Fornaciari, G.L. Mazzoli, P. Dioli, Monitoring of the invasive Halyomorpha halys, a new key pest of fruit orchards in northern Italy, J. Pest. Sci. 90 (2017) (2004) 1231–1244, https://doi.org/10.1007/s10340-017-0896-2.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Y. Lu, Y. Huang, R. Lu, Innovative Hyperspectral Imaging-Based Techniques for Quality Evaluation of Fruits and Vegetables: A Review, Applied Sciences 2017, Vol. 7, Page 189 7 (2017) 189. Doi: 10.3390/APP7020189.
[10] 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 Bioproc. Tech. 5 (2012) 1121–1142, https://doi.org/10.1007/s11947-011-0725-1.
[11] 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.
[12] B. Li, T. Ma, L. Bai, T. Inagaki, H. Seki, S. Tsuchikawa, Three-dimensional visualization and detection of early bruise in apple based on near-infrared hyperspectral imaging coupled with geometrical influence correction, Postharvest Biol. Technol. 210 (2024) 112753, https://doi.org/10.1016/j.postharvbio.2023.112753.
[13] Y. Bu, J. Luo, J. Li, Q. Chi, W. Guo, Detection of hidden bruises on kiwifruit using hyperspectral imaging combined with deep learning, Int. J. Food Sci. Technol. 59 (2024) 5975–5984, https://doi.org/10.1111/ijfs.17256.
[14] 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 10.1016/J.TIFS.2021.12.021 418 438.
[15] 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.
[16] C. Ferrari, G. Foca, R. Calvini, A. Ulrici, Fast exploration and classification of large hyperspectral image datasets for early bruise detection on apples, Chemom. Intel. Lab. Syst. 146 (2015) 108–119, https://doi.org/10.1016/j.chemolab.2015.05.016.
[17] J.M. Amigo, H. Babamoradi, S. Elcoroaristizabal, Hyperspectral image analysis, A Tutorial, Anal Chim Acta 896 (2015) 34–51, https://doi.org/10.1016/j.aca.2015.09.030.
[18] 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.
[19] 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, Journal of Spectral, Imaging (2019) a4, https://doi.org/10.1255/jsi.2019.a4.
[20] 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.
[21] D. Zhang, M.M. Islam, G. Lu, A review on automatic image annotation techniques, Pattern Recognit. 45 (2012) 346–362, https://doi.org/10.1016/j.patcog.2011.05.013.
[22] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, J. Electron. Imaging 13 (2004) 146–168, https://doi.org/10.1117/1.1631315.
[23] M. Huang, J. Tang, B. Yang, Q. Zhu, Classification of maize seeds of different years based on hyperspectral imaging and model updating, Comput. Electron. Agric. 122 (2016) 139–145, https://doi.org/10.1016/j.compag.2016.01.029.
[24] K. Esbensen, P. Geladi, Strategy of multivariate image analysis (MIA), Chemom. Intel. Lab. Syst. 7 (1989) 67–86, https://doi.org/10.1016/0169-7439(89)80112-1.
[25] J.M. Prats-Montalbán, A. De Juan, A. Ferrer, Multivariate image analysis: a review with applications, Chemom. Intel. Lab. Syst. 107 (2011) 1–23, https://doi.org/10.1016/j.chemolab.2011.03.002.
[26] M. Vidal, J.M. Amigo, Pre-processing of hyperspectral images, Essential Steps before Image Analysis, Chemometrics and Intelligent Laboratory Systems 117 (2012) 138–148, https://doi.org/10.1016/j.chemolab.2012.05.009.
[27] P.K. Sahoo, S. Soltani, A.K.C. Wong, A survey of thresholding techniques, Comput vis Graph Image Process 41 (1988) 233–260, https://doi.org/10.1016/0734-189X(88)90022-9.
[28] H.D. Cheng, X.H. Jiang, Y. Sun, J. Wang, Color image segmentation: advances and prospects, Pattern Recognit. 34 (2001) 2259–2281, https://doi.org/10.1016/S0031-3203(00)00149-7.
[29] N. Otsu, A threshold selection method from gray-level histograms, Automatica 11 (1975) 23–27.
[30] S. Munera, A. Rodríguez-Ortega, N. Aleixos, S. Cubero, J. Gómez-Sanchis, J. Blasco, Detection of Invisible Damages in 'Rojo Brillante' Persimmon Fruit at Different Stages using Hyperspectral Imaging and Chemometrics, Foods 10 (2021) 2170, https://doi.org/10.3390/foods10092170.
[31] X. Xu, S. Xu, L. Jin, E. Song, Characteristic analysis of Otsu threshold and its applications, Pattern Recognit. Lett. 32 (2011) 956–961, https://doi.org/10.1016/j.patrec.2011.01.021.
[32] T.Y. Goh S.N. Basah H. Yazid M.J. Aziz Safar F.S. Ahmad Saad Performance analysis of image thresholding: Otsu technique, Measurement 114 2018 10.1016/j.measurement.2017.09.052 298 307.
[33] J. Li, L. Chen, W. Huang, Detection of early bruises on peaches (Amygdalus persica L.) using hyperspectral imaging coupled with improved watershed segmentation algorithm, Postharvest Biol Technol 135 (2018) 104–113, https://doi.org/10.1016/j.postharvbio.2017.09.007.
[34] A. Siedliska, P. Baranowski, W. Mazurek, Classification models of bruise and cultivar detection on the basis of hyperspectral imaging data, Comput. Electron. Agric. 106 (2014) 66–74, https://doi.org/10.1016/j.compag.2014.05.012.
[35] H. Yin, B. Li, Y. Liu, F. Zhang, C. Su, A. Ou-yang, Detection of early bruises on loquat using hyperspectral imaging technology coupled with band ratio and improved Otsu method, Spectrochim. Acta A Mol. Biomol. Spectrosc. 283 (2022) 121775, https://doi.org/10.1016/j.saa.2022.121775.
[36] Y. Lu, R. Lu, Histogram-based automatic thresholding for bruise detection of apples by structured-illumination reflectance imaging, Biosyst. Eng. 160 (2017) 30–41, https://doi.org/10.1016/j.biosystemseng.2017.05.005.
[37] N. Vélez Rivera, J. Gómez-Sanchis, J. Chanona-Pérez, J.J. Carrasco, M. Millán-Giraldo, D. Lorente, S. Cubero, J. Blasco, Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning, Biosyst Eng 122 (2014) 91–98, https://doi.org/10.1016/J.BIOSYSTEMSENG.2014.03.009.
[38] L. Nørgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck, S.B. Engelsen, Interval Partial Least-Squares Regression (i PLS): a Comparative Chemometric Study with an example from Near-Infrared Spectroscopy, Appl. Spectrosc. 54 (2000) 413–419, https://doi.org/10.1366/0003702001949500.
[39] C. Ferrari, G. Foca, A. Ulrici, Handling large datasets of hyperspectral images: reducing data size without loss of useful information, Anal. Chim. Acta 802 (2013) 29–39, https://doi.org/10.1016/j.aca.2013.10.009.
[40] 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.
[41] 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.
[42] 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.), Sustainable Agriculture Reviews, Springer International Publishing, Cham, 2015: pp. 65–152. Doi: 10.1007/978-3-319-09132-7_2.
[43] 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.
[44] 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 N.K. Dhar A.K. Dutta Image Sensing Technologies: Materials, Devices, Systems, and Applications V 2018 SPIE, Orlando, United States 19 10.1117/12.2304918.
[45] R. Van Den Boomgaard, R., Van Balen, Methods for fast morphological image transforms using bitmapped binary images, CVGIP, Graphical Models and Image Processing 54 (1992) 252–258, https://doi.org/10.1016/1049-9652(92)90055-3.
[46] M. Hickey, C. King, The Cambridge Illustrated Glossary of Botanical terms, Cambridge University Press, 2000.
[47] S. Kucheryavskiy, A new approach for discrimination of objects on hyperspectral images, Chemom. Intel. Lab. Syst. 120 (2013) 126–135, https://doi.org/10.1016/j.chemolab.2012.11.009.
[48] L. Pieszczek, M. Daszykowski, Integrating hyperspectrograms with class modeling techniques for the construction of an effective expert system: Quality control of pharmaceutical tablets based on near-infrared hyperspectral imaging, J. Pharm. Biomed. Anal. 256 (2025) 116697, https://doi.org/10.1016/J.JPBA.2025.116697.
[49] P. Geladi, H.F. Grahn, Multivariate Image Analysis, in: R.A. Meyers (Ed.), Encyclopedia of Analytical Chemistry, 1st ed., Wiley, 2000. Doi: 10.1002/9780470027318.a8106.
[50] M. Cocchi, A. Biancolillo, F. Marini, Chemometric Methods for Classification and Feature selection, in, Comprehensive Analytical Chemistry, Elsevier (2018) 265–299, https://doi.org/10.1016/bs.coac.2018.08.006.
[51] D. Ballabio, F. Grisoni, R. Todeschini, Multivariate comparison of classification performance measures, Chemom. Intel. Lab. Syst. 174 (2018) 33–44, https://doi.org/10.1016/j.chemolab.2017.12.004.
[52] G. Marrubini, A. Papetti, E. Genorini, A. Ulrici, Determination of the Sugar Content in Commercial Plant Milks by Near Infrared Spectroscopy and Luff-Schoorl Total Glucose Titration, Food Anal, Methods 10 (2017) 1556–1567, https://doi.org/10.1007/s12161-016-0713-1.
[53] D.A. Burns E.W. Ciurczak Handbook of Near-Infrared Analysis 3rd Edition, 2007 CRC Press 10.1201/9781420007374.
[54] Jr. Workman Jerry, L. Weyer, Practical Guide to Interpretive Near-Infrared Spectroscopy 0 ed., 2007 CRC Press 10.1201/9781420018318.
[55] 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.
[56] 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.
[57] 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.
[58] 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.
[59] K. Kubitzki, Flowering Plants, Celastrales, Oxalidales, Rosales, Cornales, Ericales, Springer Science & Business Media, Dicotyledons, 2013.