Food Control 190 (2026) 112382

Contents lists available at ScienceDirect

Food Control

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

Use of morphological and chemical features from near infrared-hyperspectral imaging (NIR-HSI) for advanced classification of rice varieties: A chemometric approach

Elena Cazzaniga [a], Rodrigo Rocha de Oliveira [b,1], Nicola Cavallini [a,1,*], Francesco Savorani [a], Anna de Juan [b]

[a] Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 25, 10129, Torino, Italy
[b] Chemometrics Group, Universitat de Barcelona, Department of Chemical Engineering and Analytical Chemistry, Martí i Franquès, 1, 08028, Barcelona, Spain

ARTICLE INFO

Keywords:
Rice variety
Exploratory analysis
NIR imaging
Hierarchical classification
Morphogram
Morphology features

ABSTRACT

Rice is one of the world's most important staple crops, but its large varietal diversity makes authenticity control and fraud prevention difficult. This study investigated whether near-infrared hyperspectral imaging (NIR-HSI) combined with chemometrics could support rice variety discrimination through morphological and chemical information. A total of 47 rice varieties, 36 of which Italian and 11 from different foreign countries, were analysed, with 115 grains per variety, and the resulting image data were explored by principal component analysis and classified using partial least squares discriminant analysis (PLS-DA) and hierarchical modelling.

Morphological features gave the best performance: the five grain-shape classes were classified with 94.1 % NER and up to 96.0 % class accuracy, while hierarchical morphology-based models further improved discrimination, separating 11–12 groups with accuracies above 93 % and reaching 99.6 % in the best case. NIR spectra were less informative for variety discrimination: amylose-based PLS-DA achieved 86.7 % NER and 91.0 % accuracy, whereas hierarchical spectral models showed high specificity but low sensitivities for several classes. Data fusion of morphological and spectral variables did not improve the predictive performance beyond morphology alone, suggesting limited complementarity between the two information sources.

Overall, the results demonstrate that NIR-hyperspectral imaging is a rapid, non-destructive, and promising approach for rice authentication, with morphological descriptors providing the most robust basis for classification. These findings support the use of chemometric imaging strategies to strengthen quality control and help prevent food fraud in the rice supply chain.

1. Introduction

Within the big family of cereal products liable to human consumption, rice is the most consumed all over the world. According to the FAO statistics, 90 % of the global rice production comes from Asia, with China, India, and Indonesia as main producers. The rest of the global production comes from America (4.8 %), Africa (4.7 %), Europe (0.5 %) and Oceania (0.1 %). Among the European producers, Italy is the most representative country, with a share of 53 % of the total production in this continent (Giuliana et al., 2022). Italian cultivars are mostly located in the North-West territories, in particular Piedmont and Lombardy regions, which contribute to 93.2 % of national production (Arcieri & Ghinassi, 2020).

Rice is a very appreciated food, with high nutritional value due to its considerable content of carbohydrates (around 80 % (Chaudhari et al., 2018),), which represent a significant source of energy for the human body. Moreover, rice is also rich in vitamins and minerals, as well as antioxidants such as flavonoids, anthocyanins, tocopherols, etc. (Moongngarm et al., 2012). Different local cultures exhibit preferences towards specific rice varieties based on flavour, consistency, and stickiness, resulting in rice recipes that represent the hallmark of the land of origin, such as risotto, paella, and sushi (Carcea, 2021).

Rice, as a crop, has a very wide genetic diversity: thousands of different rice varieties exist (Fukagawa & Ziska, 2019), although the

* Corresponding author.
E-mail address: nicola.cavallini@polito.it (N. Cavallini).
[1] The authors equally contributed to this work.

https://doi.org/10.1016/j.foodcont.2026.112382
Received 25 February 2026; Received in revised form 9 June 2026; Accepted 10 June 2026
Available online 11 June 2026

most famous and farmed are *Oryza sativa* and *Oryza glaberrima*, mostly grown in Asia and Africa, respectively. The first way to differentiate among rice varieties is to consider structural parameters, such as the grain size, which can be short, medium, long and round. Size and shape are two qualitative attributes that describe the morphological aspect of the rice grain, together with colour and chalkiness, while chemical indicators of the rice quality are moisture and contents of proteins, sugars, and lipids (Wang et al., 2014).

One critical point in the rice industry concerns the assessment of these qualitative indicators, since this task involves analytical methods that are generally complex and time-consuming. Moreover, the classification of the different varieties is still challenging, since many varieties have similar appearance, but substantial differences in their quality attributes. This difficult distinction among varieties leads to a critical concern that characterises the modern food market: the issue of food fraud. In the rice industry, food fraud typically involves the sale of counterfeit rice, where a low-quality variety is sold as a premium product. For example, Basmati rice, with its peculiar flavour and aroma, is often adulterated with cheaper non-aromatic varieties (Śliwińska-Bartel et al., 2021).

Lots of strategies have been implemented to improve both the analytical methods in the agri-food field and the ability to fight food frauds. One that works properly in this sense is the use of vision systems and image analysis (Tillett, 1991) to perform visual inspection of meat (Elmasry et al., 2012; Kamruzzaman et al., 2012), fish (Cheng & Sun, 2014; He et al., 2015; Qin et al., 2020), fruit and vegetables (Lu et al., 2017; Pu et al., 2015), cereals (Caporaso et al., 2018; Fox & Manley, 2014; Sendin et al., 2018) and bakery products (Amigo et al., 2023), as well as plant growth monitoring and morphometry of cultivars (Brosnan & Sun, 2004; Tillett, 1991). The main advantages of this technique are the speed of measurement, cost effectiveness and accuracy, which solve the problem of human inspection characterized by high variability, subjectivity and labour costs.

A step beyond conventional vision systems is the use of hyperspectral imaging (HSI) (Khan et al., 2018). This analytical method allows obtaining two types of information about the samples: the morphological traits (which can be directly measured from the whole image) and the chemical fingerprint, since each pixel in the image is linked to a complete spectrum covering a specified wavelength range (Selci, 2019). This feature of HSI is very useful when performing analyses of sample areas, where mapping the chemical components of the sample is often of primary interest. Starting from the 1990s, Near infrared-based HSI (NIR-HSI) systems gained popularity in the agri-food industry (Manley, 2014). Lots of studies have been done exploiting HSI in the NIR range, with different purposes: ensuring food quality (Fernández Pierna et al., 2012; Mo et al., 2017; Siripatrawan & Makino, 2015), food product classification (Lei et al., 2019; Williams & Kucheryavskiy, 2016), inspection of qualitative properties of food (Barnaby et al., 2020; Calvini et al., 2020; Mulvey, 2020), and product authentication to prevent food fraud (Mendez et al., 2019). To handle the plentiful amount of information generated by this technique, NIR-HSI is often associated with the use of chemometric multivariate data analysis tools (Kumaravelu & Gopal, 2015, pp. 8–12; Nobari Moghaddam et al., 2022), which allow the extraction of information from complex and massive datasets.

More specifically and in relation to rice, HSI and different chemometric techniques have been used for characterization, classification of varieties (Fabiyi et al., 2020; Ge et al., 2024; Li et al., 2024; Onmankhong et al., 2022; Sun et al., 2021; Van De Steene et al., 2023; Wang et al., 2014; Weng et al., 2020a) and prediction of other chemical properties (Song et al., 2023; Wang et al., 2014). For instance, moisture and fatty acid content of 13 varieties from China were studied by Song et al., using HSI and different chemometric approaches (Song et al., 2023).

An interesting study from 2015 by Wang et al. as conducted on 3 Chinese rice varieties, using the spectral information form HSI but also the extracted shape features, to build classification models based on back propagation neural network, with and without fusing the two sources of information (Wang et al., 2014).

The study most similar to our work would surely be the one published in 2020 by (Fabiyi et al., 2020), where HSI and RGB images were used to gather both chemical and spatial (i.e., morphological) information. However, Fabiyi et al. studied 90 rice varieties all from Vietnam, performing classification using a Random Forest approach, trying different combinations of full and reduced datasets. Other relevant studies based on HSI and different chemometric methods are mostly from Asia and deal with Chinese or Thai rice varieties, including up to 10 varieties. Weng et al. studied 10 Chinese varieties and also combined spectra, morphological and textural features (Weng et al., 2020a), Weng et al., 2020bs well as Sun et al. (Sun et al., 2021), studying 5 Chinese rice varieties.

Considering rice varieties from Italy, a relevant study would be the one published in 2023 by Van De Steene et al., in which rice samples from 5 proveniences were analysed with many analytical techniques (from HSI to chromatography): however, in this case the precise variety of the rice samples was not reported, so a direct comparison with our results was not possible (Van De Steene et al., 2023).

In this work, we propose the application of NIR-HSI to analyse 47 different rice varieties. The information about the shape of the rice kernel, expressed as morphological features, and chemical information contained in the NIR spectra, were the seeding input to explore and classify the varieties. Firstly, Principal Component Analysis (PCA) and clustering were used as preliminary information to define classes and understand their similarity. Partial Least Squares-Discriminant Analysis (PLS-DA) classification models were built to separate the main morphological and chemical rice typologies.

Subsequently, for a finer discrimination of the large number of available rice varieties, hierarchical classification models, built based on the clustering results, and on the automatic hierarchical model builder (AHIMBU (Marchi et al., 2022),), were applied. The capability to distinguish among the studied rice varieties using both the morphological and chemical information derived from NIR-HSI and the performances of the different hierarchical schemes to build classification models are described in detail.

The main differences that can be traced between our study and the ones reported in literature would regard how deeply the Italian rice varieties have been studied using HSI, morphological features and chemometrics: in our study 36 out of 47 varieties were from Italy, while the rest was from other foreign countries. Regarding the data analysis part, we could not find any application involving hierarchical classifications, as mostly deep learning and neural network methods have been employed: these methods generally suffer from poor interpretability of the results, especially from the point of view of the chemical composition and spectral assignment of signals.

2. Materials and methods

2.1. Rice samples

Ente Nazionale Risi, a public economic body supervised by the Italian Ministry of Agriculture, Food Sovereignty and Forestry, provided rice grains samples of 47 different rice varieties: 36 Italian, 11 from all over the world, in particular Sri Lanka, Malaysia, Bangladesh, Philippines, Indonesia, Taiwan, Australia, Brazil, India, and Louisiana. All descriptive information about the samples is summarized in Table 1.

Three levels of amylose content were defined, namely “High” (>20 % amylose), “Low” (5–20 % amylose), and “Waxy” (<5 % amylose). The latter actually refers to a particular rice type, whose starch is mainly composed by amylopectin, while amylose is present just in a low percentage (Delwiche & Graybosch, 2002; Xie et al., 2022): this feature lends them a peculiar sticky consistency during cooking, which is not generally seen for all the other rice types whose amylose content is

Table 1
Information about the rice varieties provided by Ente Nazionale Risi, with specifications concerning taxonomy, origin, shape of the rice grain and amylose content.

Variety name	Taxonomy	Origin	Shape of the rice grain	Amylose content
Arborio	*Oryza sativa* (Japonica)	Italy	Long A from internal market	Low
Argo	*Oryza sativa* (Japonica)	Italy	Medium	High
Baldo	*Oryza sativa* (Japonica)	Italy	Long A from internal market	Low
Carnaroli	*Oryza sativa* (Japonica)	Italy	Long A from internal market	High
Castelmochi	*Oryza sativa* (Japonica)	Italy	Round	Waxy
CL12	*Oryza sativa* (Japonica)	Italy	Round	Low
CL15	*Oryza sativa* (Japonica)	Italy	Round	Low
CL18	*Oryza sativa* (Japonica)	Italy	Round	Low
CL26	*Oryza sativa* (Japonica)	Italy	Long B	High
CL28	*Oryza sativa* (Japonica)	Italy	Long B	High
CL31	*Oryza sativa* (Japonica)	Italy	Long A for parboil	Low
CL33	*Oryza sativa* (Japonica)	Italy	Long A for parboil	Low
CL35	*Oryza sativa* (Japonica)	Italy	Long A for parboil	Low
CL71	*Oryza sativa* (Japonica)	Italy	Long B	High
CL80	*Oryza sativa* (Japonica)	Italy	Long B	High
CL338	*Oryza sativa* (Japonica)	Italy	Long A from internal market	Low
CL510	*Oryza sativa* (Japonica)	Italy	Long A from internal market	Low
Cripto	*Oryza sativa* (Japonica)	Italy	Round	High
CRLB1	*Oryza sativa* (Japonica)	Italy	Long B	High
CRW3	*Oryza sativa* (Japonica)	Italy	Round	Waxy
Dedalo	*Oryza sativa* (Japonica)	Italy	Long B	High
Drago	*Oryza sativa* (Japonica)	Italy	Long A for parboil	Low
Duilio	*Oryza sativa* (Japonica)	Italy	Medium	Low
Elio	*Oryza sativa* (Japonica)	Italy	Round	High
Europa	*Oryza sativa* (Japonica)	Italy	Long A for parboil	Low
Iarim	*Oryza sativa* (Japonica)	Italy	Long B	High
Italmochi	*Oryza sativa* (Japonica)	Italy	Round	Waxy
Lince	*Oryza sativa* (Japonica)	Italy	Long A for parboil	Low
Padano	*Oryza sativa* (Japonica)	Italy	Medium	Low
Pegaso	*Oryza sativa* (Japonica)	Italy	Long B	High
Proteo	*Oryza sativa* (Japonica)	Italy	Round	High
Puma	*Oryza sativa* (Japonica)	Italy	Long A for parboil	Low
S. Andrea	*Oryza sativa* (Japonica)	Italy	Long A from internal market	Low
Selenio	*Oryza sativa* (Japonica)	Italy	Round	Low
Tiberio	*Oryza sativa* (Japonica)	Italy	Long A for parboil	High
Valente	*Oryza sativa* (Japonica)	Italy	Long A for parboil	Low
Kaluheenati	*Oryza sativa* (Indica)	Sri Lanka	Medium	Not defined
Hetadawee	*Oryza sativa* (Indica)	Sri Lanka	Medium	Not defined
IR6	*Oryza sativa* (Indica)	Philippines	Long B	High
IR50	*Oryza sativa* (Indica)	Philippines	Long B	Not defined
IR64	*Oryza sativa* (Indica)	Philippines	Long B	High
Cisokan	*Oryza sativa* (Indica)	Indonesia	Medium	High
Taichung Sen 17	*Oryza sativa* (Japonica)	Taiwan	Medium	Not defined
Doongara	*Oryza sativa* (Japonica)	Australia	Long B	High
IAC 165	*Oryza sativa* (Indica)	Brazil	Long A for parboil	High
Fedearroz 50	*Oryza sativa* (Indica)	South America	Long B	High
Cypress	*Oryza sativa* (Japonica)	Louisiana	Long B	High

predominant.

A visual representation of the morphological information is displayed in Fig. 1, where the five different possible shapes, i.e., “Round”, “Medium”, “Long A from internal market”, “Long A for parboil”, and “Long B” are represented.

The chemical and the morphological information reported in Table 1 was used as labelling information for exploratory analysis purposes, and as active information to define classes in classification models. Rice varieties were also used to define classes in classification models.

The rice grains were initially covered by their husk, which was manually removed before the acquisition of the hyperspectral images. Both before and after the acquisition, the samples were stored in plastic bags at room temperature (~20 °C). For each variety, a total amount of 115 rice grains was analysed.

2.2. Hyperspectral imaging instrumentation

A NIR-HSI camera (Specim FX17, Oulu, Finland) connected with a motorized scanning bed (LabScanner Setup 40 × 20 cm) was used for acquiring the images. The instrumental settings for this setup were: 935–1720 nm spectral range, 2.25 mm/s bed scanning speed, 3.5 ms

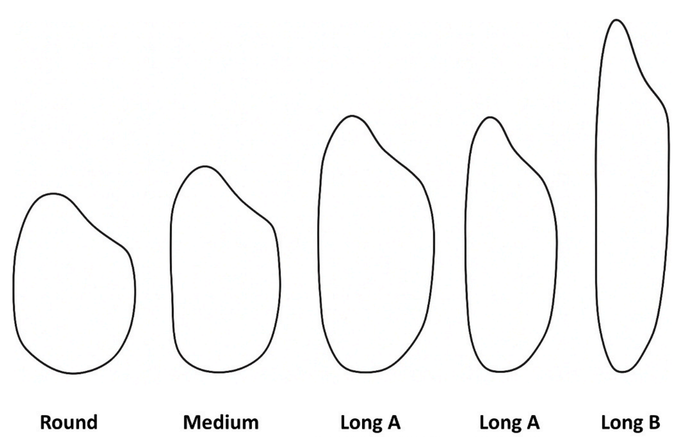

Fig. 1. Different shapes of the rice grains. Information provided by Ente Nazionale Risi.

exposure time, 32 Hz frame rate. An internal black reference standard (camera shutter) and an external white reference Spectralon surface were used for image calibration. The black standard represents the complete absence of light reflectance, while the white one provides the highest diffuse reflectance. For the instrumental settings and the image acquisition, the LUMO (Specim, Spectral Imaging LTD, Oulu, Finland) software was used.

2.2.1. Image acquisition and data treatment

For each variety, 115 grains were analysed. All measurements were performed in reflectance mode and the grains were placed in the same orientation on a black non-reflective support for measurement. Subsequently, an inhouse developed MATLAB application user interface was used to primarily preprocess the raw images by converting the reflectance signal to absorbance. Then, the images were binned to obtain squared pixels, since the vertical dimension was double than the horizontal one, due to the frame rate used for image acquisition. The final dimension of the pixel was 0.15×0.15 mm^2. All other preprocessing steps were done later in MATLAB (version 2020a, The Mathworks Inc., Natick, USA) environment using in-house made routines.

2.3. Data preprocessing

The data analysis pipeline started by preprocessing the image hypercubes, sized px, py, and λ, where px and py are the number of pixels in x and y dimensions, respectively, and λ the number of spectral channels (i.e., the acquired wavelengths) in each NIR spectrum of the image. The first preprocessing step involved the removal of the background from the images, with the aid of an in-house written MATLAB function. The image background was removed by setting a threshold of global absorbance for the pixel spectra: all absorbance values above the threshold were replaced by zeroes. This operation facilitated the removal of noisy signals due to the high absorbance of the dark background.

Afterwards, the spatial distortion caused by the camera lens was corrected to avoid misrepresenting the actual shape of the rice grains. To characterise this distortion, a precision ruler with 1 mm graduation marks was imaged across the full field of view of the sensor (640 spatial pixels). The positions of the millimetre marks were automatically detected as intensity peaks along the across-track direction using MATLAB's findpeaks function, and the midpoints between consecutive peaks provided experimental correspondences between pixel index and physical position in millimetres. A third-order polynomial was fitted to these correspondences, capturing the non-uniform pixel spacing across the field of view. The hyperspectral cube was then resampled onto an equally-spaced spatial grid by evaluating the polynomial at every pixel position and applying cubic interpolation. This correction ensures a constant pixel pitch across the field of view, so that the morphological parameters extracted from the binary masks accurately reflect the true shape and size of the scanned grains.

The preprocessed 3D hypercube was subsequently treated to extract the morphological parameters of every rice grain, and the related average spectra, as explained in the sections below. The final information obtained from each image consisted in two 2D matrices of 115 rows each, one containing the morphological features and the other the NIR average spectra of every rice grain analysed. These matrices were then inspected with a preliminary exploratory analysis, and further classification models were developed, as detailed in Section 2.7.

In-house written MATLAB functions were used for the preprocessing and the extraction of morphological rice features. The average NIR rice spectra and all exploratory and classification analyses were carried out with the MIA and PLS_Toolbox (version 9.1, Eigenvector Research Inc. WA, USA) software package.

2.4. Morphological parameters extraction

To extract the morphological information of every individual seed in each image, MATLAB's regionprops function was used. A set of morphological parameters to be computed and analysed was defined based on the purposes of the work and the potential information carried by each parameter. The chosen parameters and their description are summarized in Table 2.

2.5. NIR spectra preprocessing

To extract the NIR information from the image hypercubes, the MIA_Toolbox (version 3.1, Eigenvector Research Inc., WA, USA) was used. In each image, each rice grain is described by several pixels, each one containing a complete NIR spectrum. In the present case, an average of the spectra (and thus, of the pixels) belonging to each individual grain was computed: a single representative spectrum for each grain was obtained. As an output, each rice variety was therefore represented by 115 NIR spectra (corresponding to the number of grains per variety), which were then preprocessed with Savitzky-Golay derivative (1st derivative, 2nd polynomial order and 11 points window) and normalization (2-Norm). The spectral range was cut to remove noisy areas at the extremities, obtaining a final spectral range of 1100–1634 nm.

2.6. Exploratory data analysis

Principal component analysis (PCA (Abdi & Williams, 2010; Bro & Smilde, 2014),) was performed to explore the information carried by the data, the presence of outliers, and, in the first place, the possibility of detecting trends among the varieties. To obtain the PCA models, the 115-rows matrices of morphological parameters and NIR spectra were both averaged to obtain one individual set of morphological features and one mean spectrum for each rice variety. As a result, two 47-rows matrices containing the averaged information of the morphological and spectral properties of the rice types under examination were obtained, onto which PCA was performed. The PCA model was built using the data preprocessed as explained in the previous sections.

For an easier visualization of the PCA results, the additional information provided by Ente Nazionale Risi and reported in Table 1 was used as metadata to colour the samples in the scores plot, for both morphological and NIR properties.

Another exploratory step concerned the application of hierarchical clustering (Lee & Yang, 2009) to inspect and define natural groups of rice varieties. The intent was to exploit the information obtained from this methodology to observe how the different varieties would group from the objective point of view of their morphological and NIR features. Moreover, as a more practical aim, the clustering result could be used as a basis to build hierarchical classification models, making the

Table 2
Morphological parameters used for the structural analysis of the samples.

Parameter	Description
Area	Measured as the number of pixels contained in every rice grain, returned as a scalar.
Circularity	A quantification of the roundness or similarity of a shape to a perfect circle. $\frac{4 \cdot \pi \cdot Area}{Perimeter^2} \cdot \left(1 - \frac{0.5}{r}\right)^2$, where $r = \frac{Perimeter}{2 \cdot \pi} + 0.5$
Eccentricity	A measurement of how elongated or stretched a shape is compared to a perfect circle. $\frac{\textit{Distance between the foci of the ellipse}}{\textit{Major axis length}}$
Major axis length	Length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region, returned as a scalar.
Minor axis length	Length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region, returned as a scalar.
Perimeter	Length of the boundary of the identified sample's region, returned as a scalar.

classification task simpler.

In the present work the agglomerative clustering approach was applied, by means of Ward's method as linkage type, and Euclidean distance to calculate the distances among the samples. The first step in this sense involved the application of hierarchical clustering to the averaged matrices of all varieties, allowing to obtain groupings of different rice types according to morphological and NIR information. Dendrograms were also coloured using the morphological and NIR-related categories in Table 1 to obtain information regarding the expected groupings (as described in Table 1) and natural groupings in the data.

2.7. Classification modelling

Partial Least Squares-Discriminant Analysis (PLS-DA (Ballabio & Consonni, 2013),) was used for classification purposes. PLS-DA is one of the most used classification algorithms, with a proven effectiveness for problems involving a small number of classes (Cavallini et al., 2022; Liu et al., 2008), e.g., the five morphological types or the three amylose levels defined in Table 1. When the number of classes is very high, as is the case for the 47 rice varieties of this work, another methodology is needed, and PLS-DA becomes the underlying algorithm of hierarchical modelling, where groups of classes are separated in subsequent PLS-DA modelling steps. For this purpose, both automatic (AHIMBU (Marchi et al., 2022),) and manual (Silla & Freitas, 2010) hierarchical models were tested to classify the data.

It is important to note that for classification purposes, the morphological features and averaged spectra for the individual grains are considered, i.e., every rice variety counts with 115 grains and, therefore, models are based on (115 × 47 rice varieties = 5405 samples). The general workflow in the classification models carried out involved the split of the preprocessed samples into training and test set, which was done using the Kennard-Stone (Kennard & Stone, 1969) algorithm on the 115 samples of each variety. The test set consisted of 33 % of the total amount of samples. The training set matrix was used to build and calibrate the model, whose performances were eventually evaluated through the test set.

2.7.1. Classification with PLS-DA

To build the PLS-DA models, the same considerations made when building the PCA models were applied: the morphological data were classified according to the shape of the rice grain, while the NIR spectra were modelled considering the amylose content of the varieties. To evaluate the model performances, four parameters were considered:

1. **Specificity (Spec)**: the ability to avoid false positives in classification. It is calculated with Equation (1).
2. **Sensitivity (Sens)**: the ability to avoid false negatives in classification. It is calculated with Equation (2).
3. **Non-Error Rate (NER)**: obtained calculating the mean of the sensitivities and it represents the ability of the model to correctly classify the samples within a class. It is calculated with Equation (3).
4. **Accuracy (Acc)**: it represents an estimation of the correct classification rate, and it is calculated with Equation (4).

$$\text{Spec} = \frac{TN}{TN + FP} \tag{1}$$

$$\text{Sens} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{NER} = \frac{1}{n}\sum_{i=1}^{n} Sens_i \tag{3}$$

$$\text{Acc} = \frac{TP + TN}{n} \tag{4}$$

where TN = true negatives, FP = false positives, TP = true positives, FN = false negatives, n = n° of samples.

2.7.2. Classification with hierarchical models

Both automatic and manual hierarchical models were used to classify the morphological and spectroscopic data. This type of model works by exploiting a series of PLS-DA models, one for each classification step that occurs during the analysis, enabling to start from the totality of the samples as a big class, and finishing with all classes separated individually.

In the case of manual hierarchical models, the PLS-DA models constituting the nodes of the structure were manually calculated based on the results of the hierarchical clustering obtained during the exploratory data analysis. It should be noted that the clustering was performed on variety-level averages (one point per variety) to define the grouping structure of the tree, while the PLS-DA models at each node were calibrated exclusively on the training set samples. Since the variety-level averages represent the general morphological or spectral profile of each rice type, the tree structure obtained using all available grains was verified to be consistent with the one obtained using only the calibration set averages.

The clustering information was used to divide the samples into groups, as described in Section 2.6. On the contrary, the AHIMBU approach works by automatically calculating each PLS-DA node from the training set, after proper preliminary selection of the appropriate number of latent variables for the analysis. The training set and test set for the hierarchical models were obtained from the dataset as described in Section 2.7. These hierarchical approaches were applied when the number of classes is larger than three.

A particularly suitable use of the hierarchical classification models has been the classification of rice grains according to their varieties. In this context, each of the 47 varieties was defined as a class and it was inspected the potential of the hierarchical methods to classify the varieties, and it was interesting how the hierarchical model structure was organized. This choice was performed considering the power of these hierarchical models to perform effectively when lots of classes are involved. To evaluate the hierarchical models' performances, the same parameters described in Section 2.7.1 were used.

3. Results and discussion

3.1. Analysis of morphological data

3.1.1. Exploratory analysis of morphological data

Both morphological and NIR data were firstly analysed by PCA. As mentioned in Section 2.6, the characteristics of the 115 rice grains of every rice variety were averaged to display more easily the results. The first two PCs generally proved to be the most informative for all PCA models, with a cumulative explained variance consistently equal or above 90 %: for this reason, the following PCs were not considered in these analyses.

The PCA results of the morphological features dataset are shown in Fig. 2. Each marker in the scores plot of Fig. 2a represents one rice variety. The scores are coloured according to the rice grain shape information provided by Ente Nazionale Risi (more details in Section 2.1).

The scores plot of Fig. 2a highlights the presence of different groups of samples, which are clearly related to the shape of the rice grains. From right to left, PC1 clearly represents the evolution from “Long” varieties to “Medium” and “Round”. Indeed, five well-defined groups can be seen, with the only exception of the variety Italmochi, which belongs to the “Round” variety, but it is located among the samples of class “Medium”. The “Round” and “Long B” rice varieties lie on the opposite sides of PC1, suggesting that their morphological features are counterposed.

The location of the different morphological groups in the score plot is explained by the loadings plot of Fig. 2b, where in the negative PC1

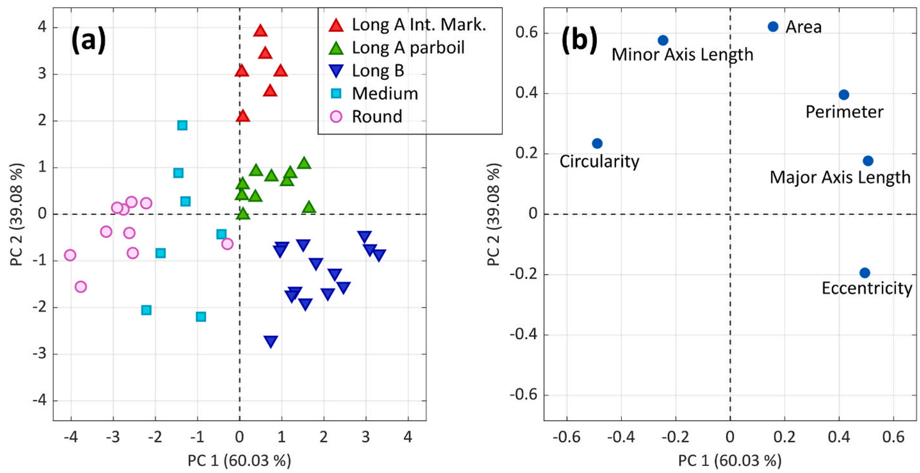

Fig. 2. PCA scores (a) and loadings (b) plots of the morphological data for the first two PCs, with colouring based on the information about the shape of the rice kernel. Each point in the scores plot represents one individual variety.

direction the Circularity can be found, which is coherent with a round shape. On the contrary, the positive PC1 direction has large loading values for Major Axis Length and Eccentricity, which are parameters showing high values for the "Long B″ varieties, as their grain shape is elongated.

To visualise and inspect the morphological parameters of the studied rice varieties we propose the "morphogram", an innovative and more detailed way to visualise together the distributions of several morphological parameters of interest in datasets with high numbers of samples. More specifically, the morphogram of an individual rice variety is elaborated by concatenating the distribution histograms of the morphological features of interest, in our case obtained from the 115 values available for each rice variety. To this aim, an in-house MATLAB routine was built.

To illustrate this tool, Fig. 3 shows the morphograms of Dedalo (in red) and Elio (in blue) rice varieties, classified as "Long B" (the most elongated rice grain type) and "Round", respectively. The figure shows six concatenated histograms (from a to f), each one corresponding to one

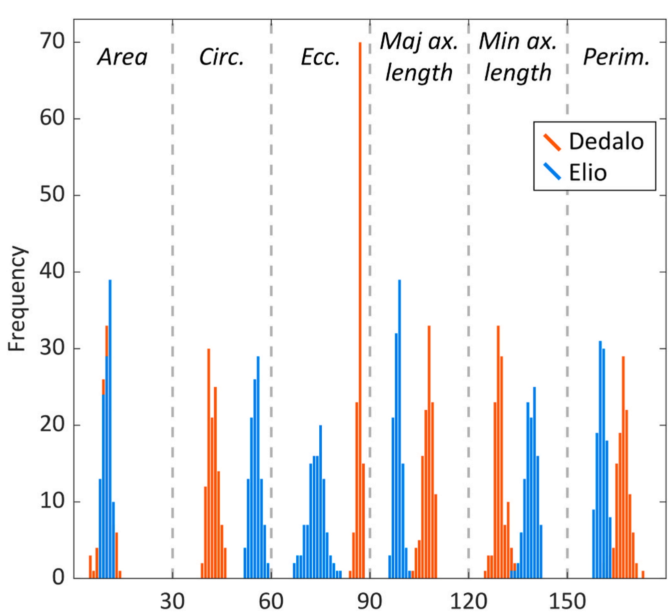

Fig. 3. A morphogram example comparing the Dedalo (Long B) and Elio (Round) varieties. The distributions of the morphological parameters are sorted from the left to the right in the following order: Area, Circularity, Eccentricity, Major axis length, Minor axis length, Perimeter.

of the morphological parameters defined in Table 1.

In this visual representation of the morphological parameters of the samples, the structural differences among them are highlighted: despite the values of the Area of the rice grains (a) are distributed very similarly for the two varieties, the other parameters underline significant differences in the shape of the kernels. The most different parameter between the two varieties under examination is Circularity (b): the two distributions are completely separated, with the round Elio at higher values. Eccentricity (c) appears antithetical with respect to Circularity, and in fact Dedalo shows higher values as its class shape is more elongated. A peculiarity of Eccentricity is that the two distributions have very different widths highlighting higher variability in this parameter for Elio (broader histogram) than Dedalo. The histograms of major axis length and minor axis length show analogous trends to eccentricity and circularity, respectively, in agreement with the expected shape features of these two rice varieties.

More than a visual representation of the morphological parameters, morphograms can represent an original way to handle and therefore model the morphological data. This version of the morphological data was analysed to look for more detailed information concerning the structural differences among the varieties under examination. Fig. 4 show the results of the PCA done using the morphograms of each rice variety. The data were preprocessed with mean centering alone.

Also in this case, the scores plot (Fig. 4a) reveals distinct groups of samples based on the rice grain shape. In particular, samples are gradually separated along PC1 according to their roundness, starting from the most elongated rice type ("Long B") on the left of PC1, and ending with the "Round" type on the right. The "Medium" type lies in a middle position between the "Long A" scores and the "Round" ones, while the "Long B" type represents the cluster that separates the most from the others, except for two samples that lie close to one "Long A for parboil" score. Analysing the loadings plot, it can be stated that the variables which mostly influence the model are Circularity and Eccentricity (second and third block of peaks). The strong negative peak in PC1 and PC2 for Eccentricity is consistent with the position of the "Long B″ scores. The "Medium" and "Round" rice types lie in the positive part of PC1, where also the loadings for the Circularity are positive and shifted to higher values for this parameter, and the "Long A" rice types lie in an average position along PC1.

Starting from this point, the development of classification models was considered to evaluate the potential of distinguishing among rice varieties. The morphogram approach proved useful for visualizing the distribution of morphological parameters across many samples, as in the case of the present work, allowing both a visual representation of the morphological parameters' distribution, and further multivariate

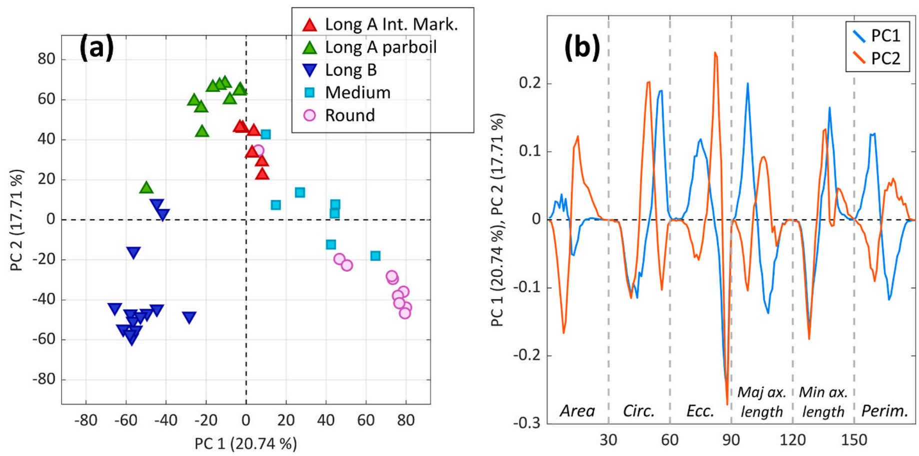

Fig. 4. PCA scores (a) and loadings (b) plots of the morphograms for the first two PCs, with colouring of the scores based on the information about the shape of the rice kernel. Each point in the scores plot represents one individual variety. In the loadings plot in (b), the distributions of the morphological parameters are sorted from the left to the right in the following order: Area, Circularity, Eccentricity, Major axis length, Minor axis length, Perimeter.

analyses. Exploratory analyses based on morphograms provided promising results, indicating them as a complementary and detailed tool for representing samples. However, since classification models are typically built per rice grain using individual morphological features (such as those employed when analysing individual grains), morphological features were taken as seeding information for hierarchical clustering.

In the optic of understanding how to classify samples by means of manual hierarchical models, a first step involved the application of clustering to obtain an idea of which groups of samples existed within the whole dataset. Thus, the clustering information was exploited to build the manual hierarchical models (Section 2.6). An example of the dendrogram obtained from the clustering for the morphological data is shown in Fig. 5.

In Fig. 5, samples are grouped according to the morphological features of the shape of the grain. This dendrogram confirmed the observation concerning the division of the samples according to the external information of shape of the rice grain: the first distinction made by the model is between the group of “Long” shapes, and the one consisting of “Medium” and “Round” shapes. Concerning the “Medium” rice type, this class shows half of its varieties close to the “Round” class, and the other half close to the “Long A from internal market” class. This observation confirms what was already seen for both PCA of morphological data and morphograms, and it is consistent with the nature of this rice type, which lies in the middle of the other two shapes (as shown in Fig. 1).

Each division was used as instruction when building the PLS-DA models that were further used in the construction of the manual hierarchical classification model to distinguish among large morphological categories and, subsequently, among rice varieties, as further explained in detail in the following section.

3.1.2. Classification models based on morphological data

To build the classification models all 115 rice grains per variety were considered, and the data were divided into calibration and test sets with the Kennard-Stone algorithm, as explained in Section 2.7. The cross-validation of all models involved the venetian blinds selection scheme with 10 splits and blind thickness equal to 1.

3.1.2.1. Classification with a single PLS-DA model. The first classification model was built using PLS-DA: the varieties were grouped into the five

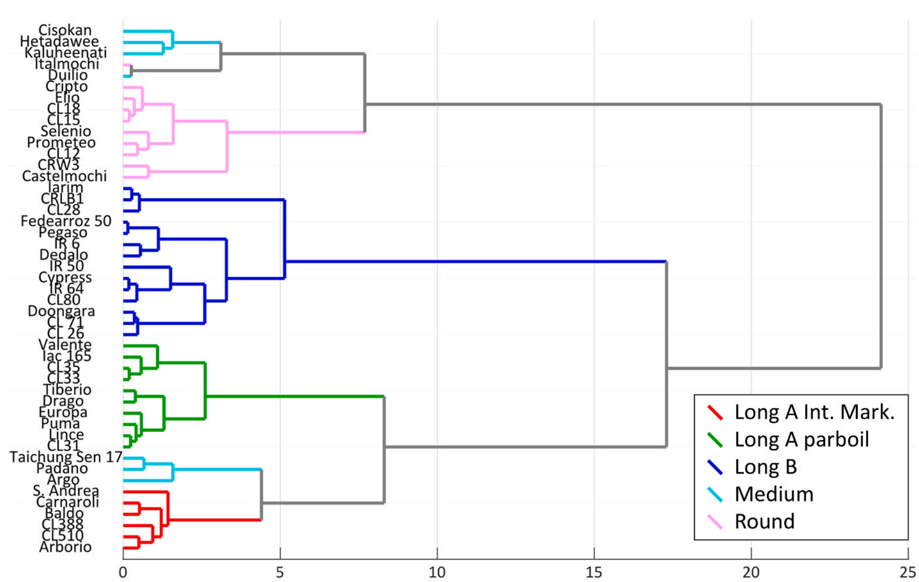

Fig. 5. Dendrogram of the hierarchical clustering applied to the morphological parameters. Rice varieties are coloured according to the shape of the rice grain as described by Ente Risi.

macro-classes defined by Ente Risi about the shape of the rice grains. The performances of this PLS-DA model are reported in Table 3.

The classification parameters suggest that the model can globally distinguish well among the five different shapes of the rice grains, as expressed by the NER parameter. The worst performances in Specificity are found for the "Medium" and "Long A for parboil" classes, with values of 67.2 % and 63.4 % respectively. This observation confirms what was already seen with the PCA model (Fig. 2), where both varieties were found in the central part of the PC1-PC2 scores plot and also in the dendrogram, where the "Medium" class is split in two parts and the "Long A for parboil" is in on the last separated branches of the dendrogram. The "Long B" class appears to be the best performing one, and this is probably due to its peculiar shape probably leading to easier recognition.

3.1.2.2. Hierarchical approaches: classification with AHIMBU. To improve the single PLS-DA model, hierarchical classification was performed. The first step involved the application of the automatic hierarchical model builder on the same dataset used for PLS-DA. The results are reported in Table 4.

For all nodes of the hierarchical classification structure PLS-DA models with 6 LVs were automatically calculated to build the AHIMBU structure, as explained in Section 2.7.2. Globally, the performances of the automatic hierarchical model are slightly better than for the individual PLS-DA model, as can be seen in the accuracy obtained in all rice classes, with a more noticeable increase for the "Medium" and "Long A for parboil" classes. In the case of the hierarchical model, the parameters concerning the Specificity proved to be better than those of PLS-DA model, while on the other hand, the single PLS-DA model shows better performances concerning the Sensitivity. However, the AHIMBU approach seems to show a more balanced quality both in terms of sensitivity and Specificity than the single PLS-DA model. Nevertheless, the differences in quality between a single PLS-DA model and AHIMBU are not high due to the small number of classes to be discriminated.

Hierarchical classification proved to be very powerful when dealing with a high number of classes, as it is the case of the 47 rice varieties in the present case. Ideally, with this approach, the model concatenates different classification models until each individual variety is separated from the others. In this work we chose to stop the class divisions when the node could not reach a threshold value of Specificity in calibration equal or greater than 80 %, assuming that this value is a practical threshold limit considering the analytical information available and the natural similarity among rice varieties. The selection of a slightly different threshold value could modify the number of groups discriminated, but never in a very significant way. The selection of a threshold of Spec(cal) > 80 % provided in all cases accuracies larger than 95 %, confirming that the threshold selected gives good and stable classification results. Using this threshold, AHIMBU modelled 11 groups of varieties, displayed in Table 5. The tree structure followed by AHIMBU in the subsequent modelling steps is reported in the Supplementary Materials as Fig. S1.

As can be seen in Table 5, AHIMBU is able to properly differentiate until 11 rice variety groups based on morphological features of grains. As the coloured information shows, it is remarkable to see that all groups are formed by rice varieties belonging to only one of the five basic categories defined in Table 1, but the hierarchical approach can go further into the morphological classification reaching 11 distinct classes. Some of the subgroups mimic well what has been observed in the dendrogram of Fig. 5. Thus, the two groups of the "Medium" class, split in separate branches of the dendrogram, are here distinguished in the same way in groups 7 and 10 and groups 6 and 7. Some other trends also match with the dendrogram structure.

Table 3
Rice kernel shape classification results with a single PLS-DA model built on morphology features. All values refer to the prediction parameters. All values except for the number of LVs are expressed in percentage.

	LVs	Sens	Spec	NER	Accuracy
Long A from internal market	4	97.9	95.3	94.1	96.0
Long A for parboil		97.9	63.4		87.3
Long B		97.1	96.3		95.0
Medium		87.9	67.2		87.2
Round		89.7	90.1		92.6

Table 4
Rice kernel shape classification results with AHIMBU performed on morphology features. All values refer to the prediction parameters. All values except for the number of LVs are expressed in percentage.

	LVs	Sens	Spec	NER	Accuracy
Long A from internal market	6	97.4	97.1	87.1	97.2
Long A for parboil		81.5	95.6		92.6
Long B		98.0	96.4		96.9
Medium		71.4	98.1		94.1
Round		87.2	97.9		95.6

Table 5
Rice variety classification results with AHIMBU performed on morphology features at the threshold of Specificity(Cal) $\geq$ 80 %. All values refer to the prediction parameters and are expressed in percentage. Rice varieties are coloured according to the five morphological categories Long A from internal market (red), Long A for parboil (green), Long B (dark blue), Medium (cyan) and Round (purple).

	Varieties	LVs	Sens	Spec	NER	Accuracy
Group 1	CL31, CL33, CL35, Drago, Duilio, Europa, IAC165, Italmochi, Lince, Puma, Tiberio	6	89.7	98.7	83.3	96.6
Group 2	CL26, CL28, CL71, CL80, CRLB1, Cypress, Dedalo, Doongara, Iarim, IR64		90.8	98.6		97.1
Group 3	Fedearroz 50, IR6, Pegaso, Valente		80.8	98.6		97.2
Group 4	CL12, CL15, CL18, Cripto, Elio, Prometeo, Selenio		84.6	99.0		96.8
Group 5	Castelmochi, CRW3		93.6	99.3		99.1
Group 6	Cisokan		82.1	98.2		97.9
Group 7	Hetadawee, Kaluheenati		84.6	98.1		97.5
Group 8	Arborio, Baldo, Carnaroli, CL388, CL510, S. Andrea		97.4	99.6		99.2
Group 9	Padano, Taichung Sen 17		84.6	98.1		97.5
Group 10	Argo		76.9	99.7		99.2
Group 11	IR50		51.3	99.3		98.3

3.1.2.3. Hierarchical approaches: classification with manual hierarchical models. Since AHIMBU provided promising results and some were similar to trends in the dendrogram, the next step involved the building of manual hierarchical models based on the tree structure provided by the dendrogram in Fig. 5. Then, individual PLS-DA models were built, where each PLS-DA here represents a "rule" (i.e., a node) of the hierarchical tree. The performances of this hierarchical model are summarized in Table 6. The number of LVs in the table is the number of LVs required in the final model that allowed the classification of every group. The classification tree followed in detail is reported in Fig. S2 of the Supplementary Materials.

The groups obtained with the manual hierarchical model were inspected as it was done with AHIMBU, and it was proven that also with the manual approach the varieties included in each group share the

Table 6
Rice variety classification results with manual hierarchical modelling performed on morphology features at the threshold of Specificity(Cal) ≥ 80 %. All values refer to the prediction parameters and are expressed in percentage. Rice varieties are coloured according to the five morphological categories Long A from internal market (red), Long A for parboil (green), Long B (dark blue), Medium (cyan) and Round (purple).

	Varieties	LVs	Sens	Spec	NER	Accuracy
Group 1	Duilio, Italmochi	4	34.6	97.8	71.6	95.1
Group 2	CL12, CL15, CL18, Cripto, Elio, Prometeo, Selenio	2	86.5	95.5		94.2
Group 3	Castelmochi, CRW3	2	96.2	98.8		98.7
Group 4	CL28, CRLB1, Iarim	1	94.0	95.3		95.3
Group 5	CL31, CL33, CL35, Drago, Europa, IAC165, Lince, Puma, Tiberio, Valente	2	88.7	95.3		93.9
Group 6	Cisokan	2	84.6	99.9		99.6
Group 7	Hetadawee, Kaluheenati	2	74.4	99.0		97.9
Group 8	Dedalo, Fedearroz 50, IR6, Pegaso	4	69.9	96.8		94.5
Group 9	CL26, CL71, CL80, Cypress, Doongara, IR50, IR64	4	65.9	98.0		93.2
Group 10	Arborio, Baldo, Carnaroli, CL388, CL510, S. Andrea	4	94.9	99.6		98.9
Group 11	Padano, Taichung Sen 17	4	7.7	99.7		95.8
Group 12	Argo	4	61.5	99.7		98.9

same type of grain shape. This highlights that also the manual model can be successfully used to exploit the information contained in the data in the identification of the samples thanks to their structural features.

By comparing the two hierarchical approaches, the first difference between the manual and the automatic hierarchical models concerns the number of groups of varieties identified by the two approaches: AHIMBU defined 11 groups, while 12 were found by the manual approach. A second difference in the performances of the manual model is that there are two groups with very low values of Sensitivity, namely Group 1 and Group 11. These groups involve rice varieties from the "Medium" group, which was shown to be the least well defined from a morphological point of view (also in PCA the "Medium" samples resulted located and distributed in middle positions in the scores space). For this reason, also the parameter of Non-Error Rate is lower than the one of AHIMBU.

Finally, the fact that both approaches (automatic and manual) provide very similar results both from the point of view of the classification performances and actual composition of the groups of variety Indicates that the morphological information that can be extracted from the data seems to be rather robust, as reproducibility seems to hold for both approaches. A choice between the two approaches can be driven by the less tedious automatic mode of AHIMBU, which allows skipping the part of the manual calculation of PLS-DA models and the building of the hierarchical structure, saving time and avoiding user-intervention mistakes. However, it should also be noted that the number of PLS-DA models based on the dendrogram approach is much lower and less computationally intensive than the exhaustive model search performed by AHIMBU.

3.2. Analysis of NIR data

3.2.1. Exploratory analysis of NIR data

The information carried by the NIR spectra was explored in the same way previously seen for morphology: after applying the preprocessing discussed in Section 2.5, the spectra of the 115 rice grains of each variety were averaged to obtain an individual representative spectrum. Then, PCA was applied to explore the information content of the data, especially concerning the amylose content of the different rice types. The PCA results are shown in Fig. 6.

PCA highlighted that the three Waxy varieties are rather well separated from the other two amylose types, which appear slightly overlapped in the central area, as it can be clearly seen in Fig. 6a. Although samples with low and high levels of amylose appear quite overlapped, there is a separation tendency along PC1, with most of the high-amylose rice varieties located on the left side of PC1, and the low-amylose samples mainly lying on the right side.

Concerning the loadings plot (Fig. 6b), a strong negative peak around 1450 nm for PC1 can be noticed, and it can be attributed to the combination of first overtone of O–H anti-symmetric and symmetric stretching of amylose (Mishra & Woltering, 2021; Han Zhang et al., 2011). This is coherent with the observations made on the scores plot, where the samples with high content of amylose lie generally on the negative side of PC1. Another strong peak can be detected along PC2, around 1400 nm, and it can be ascribed to the first overtone of O–H stretching of water (Büning-Pfaue, 2003), suggesting that the Waxy rice types could be characterized by lower content of water.

As a last exploratory step of the NIR data, a hierarchical clustering model was built. The grouping information obtained by clustering will be further applied to construct the classification models. Two cluster analyses were performed: the first one exploited the information about the amylose content of the samples, and the second one used the names of the varieties as class information. The graphical results of these analyses are reported as Fig. S3 and S4 in the Supplementary Materials.

3.2.2. Classification models based on NIR data

As discussed in Section 3.1.2 about the classification of the morphological data, the classification models based on NIR spectra were built considering all the 115 rice grains per variety. For all models, the data were divided into calibration and test set as explained in Section 2.7.

3.2.2.1. Classification with a single PLS-DA model. The information about the amylose content was used as class information for the PLS-DA model. The results are reported in Table 7.

This PLS-DA model shows high values for Sensitivity and Specificity, in particular for the Waxy varieties. This observation agrees with the PCA results, where the Waxy rice types were separated from the others in the space of the PCs. The distinction among samples with low and high levels of amylose is lower, but still acceptable.

3.2.2.2. Hierarchical approaches: classification with AHIMBU. The NIR information was used in this step to perform a hierarchical classification oriented to distinguish the 47 rice varieties. For the same reasons previously discussed, the model was stopped at the threshold of Specificity (Cal) ≥ 80 %, allowing a reliable classification of the rice varieties. The results of this classification are shown in Table 8.

From the results reported in Tables 8 and it can be noticed that the classification based on NIR information is of much lower quality with this method. In particular, the values of Sensitivity are mostly below 70 % and, consequently, the Non-Error Rate is low as well. On the other hand, the Specificity values are consistently high, all above 90 %. While this Indicates that the AHIMBU model performs well in identifying rice types that do not belong to a given class, it is important to highlight that this result might be influenced by the large number of samples analysed in this study. In the context of fighting food fraud, this model could still be useful for preliminary analyses aimed at excluding rice types that do not belong to a selected variety, but further improvements would be necessary to ensure reliable classification of samples within the target class.

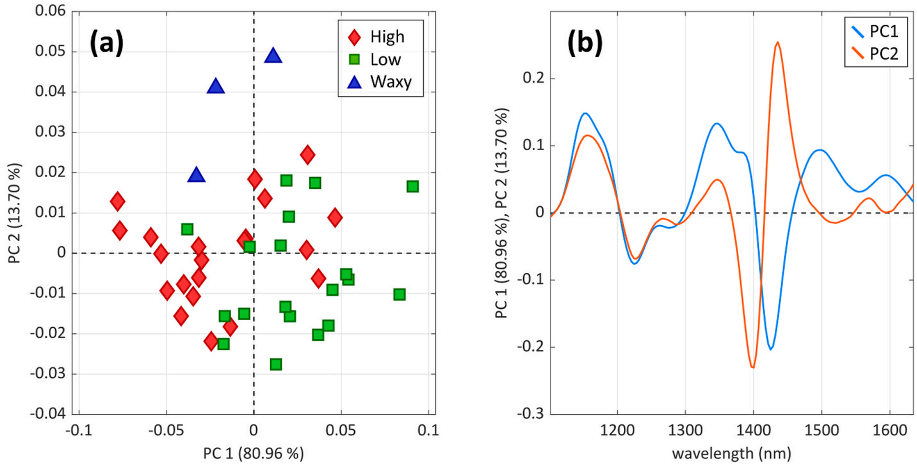

Fig. 6. PCA scores (a) and loadings (b) plots of the NIR data for the first two PCs, with colouring of the scores based on the information about the amylose content. Each point in the scores plot represents one individual variety.

Table 7
Amylose content classification results with a single PLS-DA model built on NIR spectra. All the values refer to prediction parameters. Values are expressed in percentage.

	LVs	Sens	Spec	NER	Accuracy
High	7	74.2	82.6	86.7	91.0
Low		87.7	78.5		81.6
Waxy		98.3	97.6		98.0

Based on the structure of the AHIMBU model (shown in the Supplementary Materials) it can be observed that the first four varieties that the model separates are the Italmochi, Cisokan, Castelmochi and CRW3. Apart from Cisokan, the other rice types are the three Waxy varieties present in this study. Italmochi is the first variety that is isolated from the totality of the samples as a single class itself, while Castelmochi and CRW3 are firstly grouped together and then separated one from the other. This observation underlines once again what was previously seen both in the exploratory step and in the other classification models: the Waxy rice types are detected as very different from the others. A good reason for the deficient classification is that, other than the waxy component, from a chemical point of view, rice is basically starch and all varieties have a very similar chemical composition.

3.2.2.3. Hierarchical approaches: classification with manual hierarchical models. Finally, the NIR data were classified considering each variety as a class, as seen with AHIMBU. The clustering results built in the exploratory step was exploited to choose how to group the varieties. Then, according to this division, PLS-DA models were built continuing to divide samples into classes until the threshold of Specificity ≥80 % was reached. As the performances of AHIMBU and manual model are analogous, for the sake of brevity it was chosen to detail the results of this analysis in the Supplementary Materials (Table S1). However, in this context, sensitivities never go below 10 % and have slightly higher values for the manual approach. The results of the automatic approach were shown in more detail because of the more general ease of use of this approach.

Table 8
Rice variety classification results with AHIMBU performed on NIR spectra at the threshold of Specificity(Cal) ≥ 80 %. All the values refer to the prediction parameters and are expressed in percentage.

	Varieties	LVs	Sens	Spec	NER	Accuracy
Class 1	Baldo, CL12, CL15, CL18, CL28, CL31, CL33, CL35, CL80, Cripto, CRLB1, Cypress, Drago, Duilio, Europa, Fedearroz50, IAC165, IR6, Lince, Padano, Prometeo, Puma, S. Andrea, Selenio, Taichung Sen 17, Tiberio, Valente	4	61.7	96.8	54.3	76.7
Class 2	CL26, CL,71, Dedalo, Doongara, Iarim, IR64	4	42.3	97.1		90.1
Class 3	IR50	2	92.3	91.8		91.8
Class 4	Cisokan	3	100.0	96.1		96.2
Class 5	CL510	4	48.7	98.8		97.8
Class 6	Kaluheenati	2	0.00	99.2		97.1
Class 7	Castelmochi	6	84.6	98.2		97.9
Class 8	Hetadawee	4	66.7	98.0		97.3
Class 9	Elio	4	25.6	97.9		96.4
Class 10	Pegaso	2	74.4	88.8		88.5
Class 11	Italmochi	3	97.4	96.3		96.3
Class 12	CRW3	6	2.50	99.8		97.8
Class 13	Argo	6	82.1	97.2		96.8
Class 14	Arborio	2	20.5	99.9		98.2
Class 15	CL388	2	28.2	99.1		97.7
Class 16	Carnaroli	3	41.0	99.3		98.1

3.3. Analysis of the fused dataset

The results concerning the study of morphology and NIR spectra provided in both cases interesting information about the possible differentiation among different rice varieties. While the analysis of the NIR data provided insights into rice classification, these models proved to be very limited in robustness and reliability if compared to the same kind of classification developed using the morphological features. Starting from these premises, it was decided to merge the two sources of information obtaining a fused dataset to be studied. To do that, a low-level data-fusion approach was implemented (Borràs et al., 2015; de Oliveira et al., 2020): first, the morphological and NIR data matrices were separately preprocessed as detailed in Section 2.4 and Section 2.5, then each matrix was divided by its own norm, to be finally merged, providing the fused dataset that was subject to the analyses reported in the following sections.

3.3.1. Exploratory analysis of the fused dataset

The results of the PCA on the fused data are reported in Fig. 7, where the scores are coloured based on the rice grain shape (Fig. 7a, same colours as seen for the PCA of morphology in Fig. 2a), and also based on the amylose content of each variety (Fig. 7c).

The loadings plot in Fig. 7b shows that the morphological parameters (at the left side of the plot and zoomed in Fig. 7d) strongly influence the PCA model, which is seen by the general much higher value of loadings for this kind of features. There is a certain relationship between morphological and spectroscopic features that can be deduced by observing the loadings plot. Eccentricity and Major Axis Length lie in the positive direction of PC1, as well as two NIR features, namely the absorption bands around 1225 nm and around 1426 nm, ascribable to the second overtone symmetric stretching of methyl group CH_3 of carbohydrates and fat (Yang et al., 2021; Hailiang Zhang et al., 2020), and the second overtone stretching of the O–H group of amylose, respectively. Circularity and Minor Axis Length lie in the opposite direction of PC1, together with the absorption bands around at 1150 nm and around 1345 nm, the first one referring to the absorption of carbohydrates and fat, and the second one to the first overtone of O–H stretching of water. This correlation between structural and spectroscopic features might suggest that the rice varieties with an elongated shape contains high levels of amylose, while the circular shape is more linked to high levels of water in the grain.

Looking at the scores plot, PC1 separates mainly the "Long B" rice types (positive direction of PC1) from all the rest of rice morphology types. This could suggest that varieties from "Long B″ present high levels of amylose. All the rest of morphological classes may be characterized as having low amylose content or being from the waxy type. PC2 shows even more clearly the prevalence of morphological features towards NIR, which has comparatively even lower loading values in this case. Indeed, circularity shows a high positive value, whereas major axis length and eccentricity point towards negative values. This is the reason why "Round" rice varieties lie in the upper part of the score plot (positive PC2 values) and the rest of varieties move down this axis following the sequence "Medium" and "Long" categories, all of them displaying negative PC2 values. Along the PC2 axis, as expected from the low NIR loading values, there is no structured trend between rice varieties with low, high amylose contents or waxy, meaning that the relationship between morphological features and chemical composition is very limited.

This lack of general relation between shape and NIR information has made that the classification models (not shown) performed with the fused data were not outperforming those obtained only with the morphological features. In this case, the fusion approach may help at the most to increase the exploratory side of this study but does not change

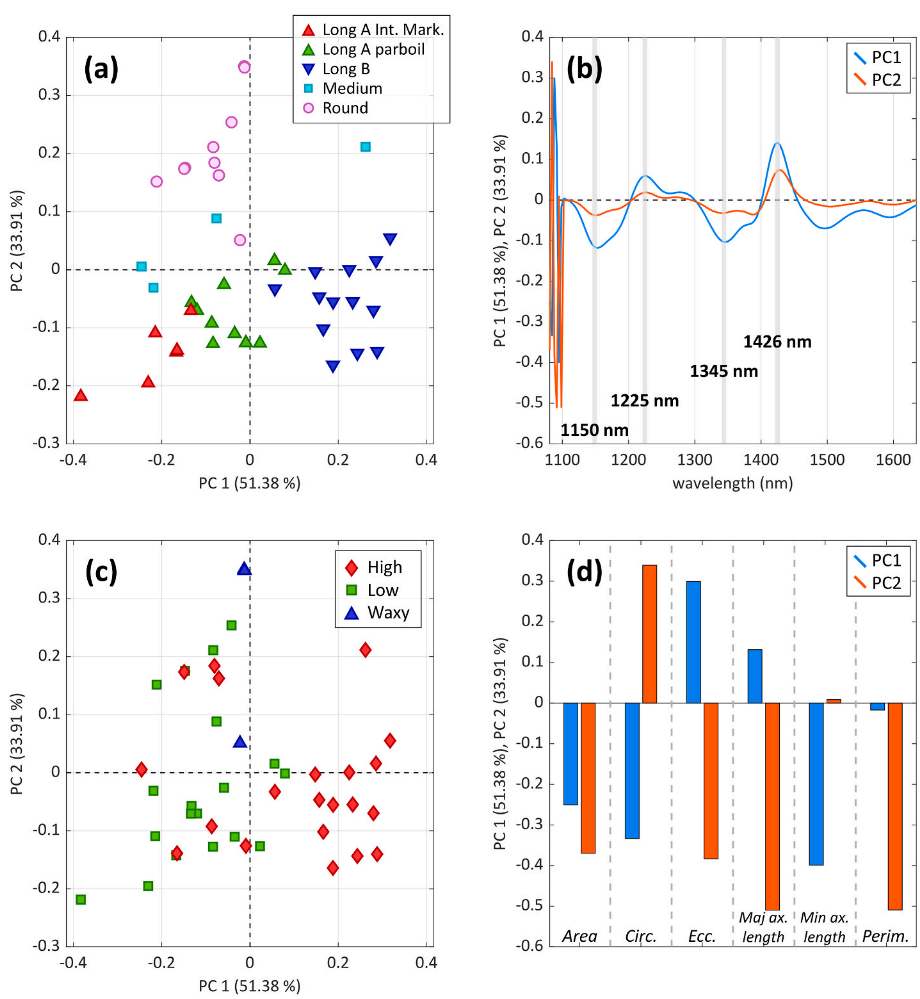

Fig. 7. PCA scores (a,c) and loadings (b,d) plots of the fused data (NIR and morphological data). The scores in (a) are coloured according to the shape of the rice grain, while the scores in (c) are coloured according to the amylose content. The plot in (d) is a magnification of the loadings corresponding to the morphological parameters, which are located at the far-right side in (b).

the final quality of the best classification models previously obtained.

3.4. Overall results discussion and contextualization

The results appear coherent with what is generally reported by previous studies (Ge et al., 2024; Li et al., 2024; Onmankhong et al., 2022; Sun et al., 2021; Van De Steene et al., 2023; Wang et al., 2014; Weng et al., 2020b), for instance the large study on 90 Vietnamese varieties by Fabiyi et al. (Fabiyi et al., 2020). Direct and punctual comparison of performances with these studies is difficult, as very different rice varieties were used: in our study many Italian varieties are present, as opposed to the large majority of the literature studies focusing on Asian varieties.

Moreover, the chemometric classification techniques are very different: we employed standard PLS-DA approaches with different levels of complexity, from a single model to hierarchical models, while most of the previous research used deep learning and neural network approaches. Also, in our case the NIR spectra are used and modelled in full, without performing any variable selection.

4. Conclusions

Food fraud is a concern with universal impacts, as it undermines the confidence in the food industry not only of specific groups of consumers, but its effect is extended to the global population. To fight this issue, rapid and precise analytical methods are needed. This work evaluated the application of hyperspectral imaging in the NIR range to analyse 47 rice varieties, inspecting the dataset information and using it to build robust and reliable classification models. The main advantage of using HSI is the possibility to simultaneously obtain spatial and chemical information, thus allowing the retrieval of morphological features and chemical information on the rice grains.

The information contained in the images, extracted and explored through multivariate analysis tools, allowed building classification models aimed at distinguishing samples based on their morphological and spectroscopic features. Three classification modelling procedures were performed: single PLS-DA classification models, more oriented to distinguish broad morphological and amylose content-based composition classes; manual and automatic hierarchical approaches, able to handle high numbers of classes, basically focused on the distinction among rice varieties. The PLS-DA models proved to be robust and effective to distinguish the broadest morphological and amylose content-based composition classes, with good sensitivity and specificity results. The best results were found for the morphology-based classification, while the NIR-based classification model had slightly lower classification parameters, but remaining robust for prediction purposes.

The hierarchical models proved to be adequate when distinction among rice varieties is the main goal. When using morphological features, the hierarchical approaches enabled a more detailed distinction than the one provided by the five broad morphological classes, reaching the possibility to classify correctly between 11 and 12 rice variety classes with good values of sensitivity and specificity. The performance using NIR spectroscopic information alone was clearly poorer, but still the discrimination between the three large classes linked to amylose content was improved, even if with some classes characterized by unusually low sensitivity values. Such an outcome may be attributable to the very similar composition among rice varieties. Both manual dendrogram-based and AHIMBU hierarchical classification approaches performed similarly in all the performed classification procedures, being AHIMBU more recommended due to its automatic modus operandi.

Data fusion of morphology and NIR spectra was studied through exploratory analysis and proved that the morphological features only had a limited relationship with the chemical information related to the amylose content, showing only a clear tendency for the variety "Long B", which seems to generally show a higher amylose content. The lack of clear complementary trends among the two kinds of information and the general similarity in terms of chemical composition in the different rice varieties explains the fact that classification models based on the fused information do not outperform those given by morphological features alone.

Given the difficulty in discriminating the rice varieties based on the NIR information and with the idea to go a step further in the knowledge of similarities and differences among rice varieties, a future line of research to be explored would be performing a one-class classification approach per each rice variety. The results of these models would provide complementary information on which varieties could be clearly differentiated from a particular one and which ones would be more easily mistaken.

CRediT authorship contribution statement

Elena Cazzaniga: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rodrigo Rocha de Oliveira:** Writing – review & editing, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nicola Cavallini:** Writing – review & editing, Visualization, Validation, Supervision, Software, Investigation, Formal analysis, Conceptualization. **Francesco Savorani:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Anna de Juan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

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

R.R.O. and A.J. acknowledge grant PID2023-146465NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2026.112382.

Data availability

Data will be made available on request.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433–459. https://doi.org/10.1002/wics.101

Amigo, J. M., Jespersen, B. M., van den Berg, F., Jensen, J. J., & Engelsen, S. B. (2023). Batch-wise versus continuous dough mixing of Danish butter cookies. A near infrared hyperspectral imaging study. *Food Chemistry, 414*, Article 135731. https://doi.org/10.1016/J.FOODCHEM.2023.135731

Arcieri, M., & Ghinassi, G. (2020). Rice cultivation in Italy under the threat of climatic change: Trends, technologies and research gaps. *Irrigation and Drainage, 69*(4), 517–530. https://doi.org/10.1002/IRD.2472

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods, 5*(16), 3790–3798. https://doi.org/10.1039/c3ay40582f

Barnaby, J. Y., Huggins, T. D., Lee, H., McClung, A. M., Pinson, S. R. M., Oh, M., Bauchan, G. R., Tarpley, L., Lee, K., Kim, M. S., & Edwards, J. D. (2020). Vis/NIR hyperspectral imaging distinguishes sub-population, production environment, and physicochemical grain properties in rice. *Scientific Reports, 10*(1), 9284. https://doi.org/10.1038/s41598-020-65999-7

Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., & Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment – A review. *Analytica Chimica Acta, 891*, 1–14. https://doi.org/10.1016/J.ACA.2015.04.042

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods, 6*(9), 2812–2831. https://doi.org/10.1039/C3AY41907J

Brosnan, T., & Sun, D. W. (2004). Improving quality inspection of food products by computer vision—a review. *Journal of Food Engineering, 61*(1), 3–16. https://doi.org/10.1016/S0260-8774(03)00183-3

Büning-Pfaue, H. (2003). Analysis of water in food by near infrared spectroscopy. *Food Chemistry, 82*(1), 107–115. https://doi.org/10.1016/S0308-8146(02)00583-6

Calvini, R., Michelini, S., Pizzamiglio, V., Foca, G., & Ulrici, A. (2020). Exploring the potential of NIR hyperspectral imaging for automated quantification of rind amount in grated Parmigiano Reggiano cheese. *Food Control, 112*, Article 107111. https://doi.org/10.1016/J.FOODCONT.2020.107111

Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2018). Near-infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. *Applied Spectroscopy Reviews, 53*(8), 667–687. https://doi.org/10.1080/05704928.2018.1425214

Carcea, M. (2021). Value of wholegrain rice in a healthy human nutrition. *Agriculture, 11* (8). https://doi.org/10.3390/AGRICULTURE11080720

Cavallini, N., Pennisi, F., Giraudo, A., Pezzolato, M., Esposito, G., Gavoci, G., Magnani, L., Pianezzola, A., Geobaldo, F., Savorani, F., & Bozzetta, E. (2022). Chemometric differentiation of sole and plaice fish fillets using three near-infrared instruments. *Foods, 11*(11), 1643. https://doi.org/10.3390/foods11111643

Chaudhari, P. R., Tamrakar, N., Singh, L., Tandon, A., Sharma, D., & Prabha Chaudhari, C. R. (2018). Rice nutritional and medicinal properties: A review article. *Journal of Pharmacognosy and Phytochemistry, 7*(2).

Cheng, J. H., & Sun, D. W. (2014). Hyperspectral imaging as an effective tool for quality analysis and control of fish and other seafoods: Current research and potential applications. *Trends in Food Science & Technology, 37*(2), 78–91. https://doi.org/10.1016/J.TIFS.2014.03.006

de Oliveira, R. R., Avila, C., Bourne, R., Muller, F., & de Juan, A. (2020). Data fusion strategies to combine sensor and multivariate model outputs for multivariate statistical process control. *Analytical and Bioanalytical Chemistry, 412*(9), 2151–2163. https://doi.org/10.1007/S00216-020-02404-2

Delwiche, S. R., & Graybosch, R. A. (2002). Identification of waxy wheat by near-infrared reflectance spectroscopy. *Journal of Cereal Science, 35*(1), 29–38. https://doi.org/10.1006/JCRS.2001.0400

Elmasry, G., Barbin, D. F., Sun, D. W., & Allen, P. (2012). Meat quality evaluation by hyperspectral imaging technique: An overview. *Critical Reviews in Food Science and Nutrition, 52*(8), 689–711. https://doi.org/10.1080/10408398.2010.507908

Fabiyi, S. D., Vu, H., Tachtatzis, C., Murray, P., Harle, D., Dao, T. K., Andonovic, I., Ren, J., & Marshall, S. (2020). Varietal classification of rice seeds using RGB and hyperspectral images. *IEEE Access, 8*, 22493–22505. https://doi.org/10.1109/ACCESS.2020.2969847

Fernández Pierna, J. A., Vermeulen, P., Amand, O., Tossens, A., Dardenne, P., & Baeten, V. (2012). NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed. *Chemometrics and Intelligent Laboratory Systems, 117*, 233–239. https://doi.org/10.1016/J.CHEMOLAB.2012.02.004

Fox, G., & Manley, M. (2014). Applications of single kernel conventional and hyperspectral imaging near infrared spectroscopy in cereals. *Journal of the Science of Food and Agriculture, 94*(2), 174–179. https://doi.org/10.1002/JSFA.6367

Fukagawa, N. K., & Ziska, L. H. (2019). Rice: Importance for global nutrition. *Journal of Nutritional Science and Vitaminology, 65*(Supplement), S2–S3. https://doi.org/10.3177/JNSV.65.S2

Ge, Y., Song, S., Yu, S., Zhang, X., & Li, X. (2024). Rice seed classification by hyperspectral imaging system: A real-world dataset and a credible algorithm. *Computers and Electronics in Agriculture, 219*, Article 108776. https://doi.org/10.1016/J.COMPAG.2024.108776

Giuliana, V., Lucia, M., Marco, R., & Simone, V. (2022). Environmental life cycle assessment of rice production in northern Italy: A case study from Vercelli. *International Journal of Life Cycle Assessment*, (8), 1523–1540. https://doi.org/10.1007/S11367-022-02109-X, 2022 29:8, 29.

He, H. J., Wu, D., & Sun, D. W. (2015). Nondestructive spectroscopic and imaging techniques for quality evaluation and assessment of fish and fish products. *Critical Reviews in Food Science and Nutrition, 55*(6), 864–886. https://doi.org/10.1080/10408398.2012.746638

Kamruzzaman, M., ElMasry, G., Sun, D. W., & Allen, P. (2012). Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. *Analytica Chimica Acta, 714*, 57–67. https://doi.org/10.1016/J.ACA.2011.11.037

Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics, 11*(1), 137–148. https://doi.org/10.1080/00401706.1969.10490666

Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K., & Abbas, A. (2018). Modern trends in hyperspectral image analysis: A review. *IEEE Access, 6*, 14118–14129. https://doi.org/10.1109/ACCESS.2018.2812999

Kumaravelu, C., & Gopal, A. (2015). A review on the applications of near-Infrared spectrometer and Chemometrics for the agro-food processing industries. *Proceedings - 2015 IEEE international conference on technological innovations in ICT for agriculture and rural development, TIAR 2015*. https://doi.org/10.1109/TIAR.2015.7358523

Lee, I., & Yang, J. (2009). Common clustering algorithms. In *Comprehensive chemometrics* (pp. 577–618). Elsevier. https://doi.org/10.1016/B978-044452701-1.00064-8.

Lei, T., Lin, X. H., & Sun, D. W. (2019). Rapid classification of commercial Cheddar cheeses from different brands using PLSDA, LDA and SPA–LDA models built by hyperspectral data. *Journal of Food Measurement and Characterization, 13*(4), 3119–3129. https://doi.org/10.1007/S11694-019-00234-0/TABLES/5

Li, S., Sun, L., Tian, Y., Lu, X., Fu, Z., Lv, G., Zhang, L., Xu, Y., & Che, W. (2024). Research on non-destructive identification technology of rice varieties based on HSI and GBDT. *Infrared Physics & Technology, 142*, Article 105511. https://doi.org/10.1016/J.INFRARED.2024.105511

Liu, L., Cozzolino, D., Cynkar, W. U., Dambergs, R. G., Janik, L., O'Neill, B. K., Colby, C. B., & Gishen, M. (2008). Preliminary study on the application of visible–near infrared spectroscopy and chemometrics to classify riesling wines from different countries. *Food Chemistry, 106*(2), 781–786. https://doi.org/10.1016/J.FOODCHEM.2007.06.015

Lu, Y., Huang, Y., & Lu, R. (2017). Innovative hyperspectral imaging-based techniques for quality evaluation of fruits and vegetables: A review. *Applied Sciences, 7*(2). https://doi.org/10.3390/APP7020189

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews, 43*(24), 8200–8214. https://doi.org/10.1039/C4CS00062E

Marchi, L., Krylov, I., Roginski, R. T., Wise, B., Di Donato, F., Nieto-Ortega, S., Pereira, J. F. Q., & Bro, R. (2022). Automatic hierarchical model builder. *Journal of Chemometrics, 36*(12). https://doi.org/10.1002/CEM.3455. Article e3455.

Mendez, J., Mendoza, L., Tirado, J. P. C., Quevedo, R. A., & Jara, R. S. (2019). Trends in application of NIR and hyperspectral imaging for food authentication. *Scientia Agropecuaria, 10*(1), 143–161. https://doi.org/10.17268/sci.agropecu.2018.01.16

Mishra, P., & Woltering, E. J. (2021). Identifying key wavenumbers that improve prediction of amylose in rice samples utilizing advanced wavenumber selection techniques. *Talanta, 224*, Article 121908. https://doi.org/10.1016/J.TALANTA.2020.121908

Mo, C., Kim, G., Kim, M. S., Lim, J., Lee, S. H., Lee, H. S., & Cho, B. K. (2017). Discrimination methods for biological contaminants in fresh-cut lettuce based on VNIR and NIR hyperspectral imaging. *Infrared Physics & Technology, 85*, 1–12. https://doi.org/10.1016/J.INFRARED.2017.05.003

Moongngarm, A., Daomukda, N., & Khumpika, S. (2012). Chemical compositions, phytochemicals, and antioxidant capacity of rice bran, rice bran layer, and rice germ. *APCBEE Procedia, 2*, 73–79. https://doi.org/10.1016/J.APCBEE.2012.06.014

Mulvey, B. W. (2020). Determination of fat content in foods using a near-infrared spectroscopy sensor. *Proceedings of IEEE sensors, 2020-October*. https://doi.org/10.1109/SENSORS47125.2020.9278647

Nobari Moghaddam, H., Tamiji, Z., Akbari Lakeh, M., Khoshayand, M. R., & Haji Mahmoodi, M. (2022). Multivariate analysis of food fraud: A review of NIR based instruments in tandem with chemometrics. *Journal of Food Composition and Analysis, 107*, Article 104343. https://doi.org/10.1016/J.JFCA.2021.104343

Onmankhong, J., Ma, T., Inagaki, T., Sirisomboon, P., & Tsuchikawa, S. (2022). Cognitive spectroscopy for the classification of rice varieties: A comparison of machine learning and deep learning approaches in analysing long-wave near-infrared hyperspectral images of brown and milled samples. *Infrared Physics & Technology, 123*, Article 104100. https://doi.org/10.1016/J.INFRARED.2022.104100

Pu, Y. Y., Feng, Y. Z., & Sun, D. W. (2015). Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: A review. *Comprehensive Reviews in Food Science and Food Safety, 14*(2), 176–188. https://doi.org/10.1111/1541-4337.12123

Qin, J., Vasefi, F., Hellberg, R. S., Akhbardeh, A., Isaacs, R. B., Yilmaz, A. G., Hwang, C., Baek, I., Schmidt, W. F., & Kim, M. S. (2020). Detection of fish fillet substitution and mislabeling using multimode hyperspectral imaging techniques. *Food Control, 114*, Article 107234. https://doi.org/10.1016/j.foodcont.2020.107234

Selci, S. (2019). The future of hyperspectral imaging. *Journal of Imaging, 5*(11). https://doi.org/10.3390/JIMAGING5110084

Sendin, K., Williams, P. J., & Manley, M. (2018). Near infrared hyperspectral imaging in quality and safety evaluation of cereals. *Critical Reviews in Food Science and Nutrition, 58*(4), 575–590. https://doi.org/10.1080/10408398.2016.1205548

Silla, C. N., & Freitas, A. A. (2010). A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery, 22*(1), 31–72. https://doi.org/10.1007/S10618-010-0175-9

Siripatrawan, U., & Makino, Y. (2015). Monitoring fungal growth on brown rice grains using rapid and non-destructive hyperspectral imaging. *International Journal of Food Microbiology, 199*, 93–100. https://doi.org/10.1016/J.IJFOODMICRO.2015.01.001

Śliwińska-Bartel, M., Burns, D. T., & Elliott, C. (2021). Rice fraud a global problem: A review of analytical tools to detect species, country of origin and adulterations. *Trends in Food Science & Technology, 116*, 36–46. https://doi.org/10.1016/J.TIFS.2021.06.042

Song, Y., Cao, S., Chu, X., Zhou, Y., Xu, Y., Sun, T., Zhou, G., & Liu, X. (2023). Non-destructive detection of moisture and fatty acid content in rice using hyperspectral imaging and chemometrics. *Journal of Food Composition and Analysis, 121*, Article 105397. https://doi.org/10.1016/J.JFCA.2023.105397

Sun, J., Zhang, L., Zhou, X., Yao, K., Tian, Y., & Nirere, A. (2021). A method of information fusion for identification of rice seed varieties based on hyperspectral imaging technology. *Journal of Food Process Engineering, 44*(9). https://doi.org/10.1111/JFPE.13797. Article e13797.

Tillett, R. D. (1991). Image analysis for agricultural processes: A review of potential opportunities. *Journal of Agricultural Engineering Research, 50*(C), 247–258. https://doi.org/10.1016/S0021-8634(05)80018-6

Van De Steene, J., Ruyssinck, J., Fernandez-Pierna, J. A., Vandermeersch, L., Maes, A., Van Langenhove, H., Walgraeve, C., Demeestere, K., De Meulenaer, B., Jacxsens, L., & Miserez, B. (2023). Fingerprinting methods for origin and variety assessment of rice: Development, validation and data fusion experiments. *Food Control, 151*, Article 109780. https://doi.org/10.1016/J.FOODCONT.2023.109780

Wang, L., Liu, D., Pu, H., Sun, D. W., Gao, W., & Xiong, Z. (2014). Use of hyperspectral imaging to discriminate the variety and quality of rice. *Food Analytical Methods, 8*(2), 515–523. https://doi.org/10.1007/S12161-014-9916-5

Weng, S., Tang, P., Yuan, H., Guo, B., Yu, S., Huang, L., & Xu, C. (2020a). Hyperspectral imaging for accurate determination of rice variety using a deep learning network with multi-feature fusion. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 234*, Article 118237. https://doi.org/10.1016/J.SAA.2020.118237

Weng, S., Tang, P., Yuan, H., Guo, B., Yu, S., Huang, L., & Xu, C. (2020b). Hyperspectral imaging for accurate determination of rice variety using a deep learning network with multi-feature fusion. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 234*, Article 118237. https://doi.org/10.1016/J.SAA.2020.118237

Williams, P. J., & Kucheryavskiy, S. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry, 209*, 131–138. https://doi.org/10.1016/J.FOODCHEM.2016.04.044

Xie, L. H., Tang, S. Q., Wei, X. J., Sheng, Z. H., Shao, G. N., Jiao, G. A., Hu, S. K., Wang-Lin, & Hu, P. S. (2022). Simultaneous determination of apparent amylose, amylose and amylopectin content and classification of waxy rice using near-infrared spectroscopy (NIRS). *Food Chemistry, 388*, Article 132944. https://doi.org/10.1016/J.FOODCHEM.2022.132944

Yang, L., Gao, H., Meng, L., Fu, X., Du, X., Wu, D., & Huang, L. (2021). Nondestructive measurement of pectin polysaccharides using hyperspectral imaging in mulberry fruit. *Food Chemistry, 334*, Article 127614. https://doi.org/10.1016/J.FOODCHEM.2020.127614

Zhang, H., Miao, Y., Takahashi, H., & Chen, J. Y. (2011). Amylose analysis of rice flour using near-infrared spectroscopy with particle size compensation. *Food Science and Technology Research, 17*(4), 361–367. https://doi.org/10.3136/FSTR.17.361

Zhang, H., Zhang, S., Chen, Y., Luo, W., Huang, Y., Tao, D., Zhan, B., & Liu, X. (2020). Non-destructive determination of fat and moisture contents in Salmon (Salmo salar) fillets using near-infrared hyperspectral imaging coupled with spectral and textural features. *Journal of Food Composition and Analysis, 92*, Article 103567. https://doi.org/10.1016/J.JFCA.2020.103567