


RESEARCH

Open Access



# Portable optical spectroscopy and machine learning techniques for quantification of the biochemical content of raw food materials

Cosimo Ricci<sup>1</sup>, Agata Gadaleta<sup>2\*</sup>, Annamaria Gerardino<sup>1</sup>, Angelo Didonna<sup>2</sup>, Giuseppe Ferrara<sup>2\*</sup>  and Francesca Romana Bertani<sup>1</sup>

## Abstract

**Background** Accuracy in determining food authenticity, possible contamination, content analysis, and even geographical origin is of considerable scientific and economic value. The aim of this study is to facilitate quantitative evaluation of protein content in the seeds of cereals (*Triticum turgidum* var. durum and *Tritordeum* genotypes) and ripening pomegranate fruits (Wonderful cultivar).

**Methods** Two species of wheat were evaluated in this study: durum wheat, *Triticum turgidum* var. durum, and *Tritordeum* (durum wheat × wild barley) together with pomegranate fruits of the variety Wonderful. Two different portable Near InfraRed (NIR) spectrometers have been used: a prototype developed in the PhasmaFood project and the commercial SCiO™ molecular sensor.

**Results** Considering the specific samples, the obtained results of the classification models indicate a validation mean absolute error of 0.8% (percentage of total protein content in dry matter) for two species of wheat using Convolutional Neural Network following normalization procedures and 0.32% using Partial Least Square (PLS) analysis applied to *Tritordeum* samples; visible reflectance spectra have been used to discriminate the two cereal species. A Root Mean Square Error (RMSE) of 1.25 was obtained for the determination of total soluble solids (TSS) over a 2-year period for pomegranate fresh fruits of Wonderful cultivar, which is commonly harvested with TSS values of 16–17.

**Conclusions** The application of portable sensors using NIR spectroscopy can be a valid and rapid alternative to the use of destructive laboratory techniques for the assessment of protein content in intact wheat seeds and ripeness grade (TSS) in intact pomegranates.

**Keywords** Portable spectroscopy, NIR spectroscopy, Chemometrics, Wheat, Pomegranate

\*Correspondence:

Agata Gadaleta  
agata.gadaleta@uniba.it

Giuseppe Ferrara  
giuseppe.ferrara@uniba.it

<sup>1</sup> CNR – IFN Institute of Photonics and Nanotechnologies, Via Fosso del Cavaliere 100, 00133 Rome, Italy

<sup>2</sup> Department of Soil, Plant and Food Science (Di.S.S.P.A.), University of Bari Aldo Moro, Bari, Italy

## Introduction

The development of rapid fingerprinting methods for the determination of crop quality is generating great interest in agricultural industries, as these are increasingly reliable tools in food analytics. For this purpose, several non-destructive detection techniques are currently available, including optical spectroscopic techniques such as visible and near-infrared (VNIR) spectroscopy, multi- and



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

hyperspectral imaging, fluorescence spectroscopy (Bertani et al. 2020) and imaging (Bertani et al. 2024), visible imaging, and colorimetry (Žilić et al. 2011; Ferrara et al. 2022a, 2022b; Li et al. 2018). Near-infrared region (NIR) applications were demonstrated to be valid tools for the determination of vitamins, water, proteins and carbohydrates, molecules that contain a wide spectrum of NIR-active chemical groups (Osborne 2006). Recent studies have demonstrated the ability of NIR to estimate the quality of fruit in the field (Donis-González et al. 2020a) and post-harvest (McGlone et al. 2017). More generally, the development of portable sensors (Müller-Maatsch et al. 2021) has opened up the possibility of taking lab analysis directly in the field or to production lines to monitor food matrices in real time and/or optimize production processes. The challenge, in this framework, is to achieve high analytical accuracy in classification/regression tasks starting with data that typically have lower spectral resolution and lower signal to noise ratios than data generated by expensive benchtop instrumentation. Moreover, seeds and single fruits are inherently less homogeneous than flour and juices thus increasing the variance in the measurements.

Analytical information is extracted from the sample spectra by application of advanced chemometrics models (Zhong and Wang 2019; Zhang et al. 2019a). In this field, a great deal of progress has been possible thanks to the deployment of machine learning (including deep learning, like in Nallan et al., 2022; Zhou et al. 2022) methods in the analysis workflow and the increasing availability of data to build reliable databases. In this study, two different food matrices and two different applications have been investigated to assess the power of the combination of portable spectroscopy and chemometrics: cereal seeds to evaluate total protein content, and pomegranate fruits to evaluate ripeness grade.

Among staple crops, cereals such as rice, wheat and maize play a key role in human nutrition and amongst cereals, durum wheat is the leading commercial crop as its consumption is the highest within Mediterranean countries. The current conflict in Ukraine is also putting pressure on the production of wheat since flour is consumed in many countries from Africa to South America.

In cereals, and more specifically in wheat, the nutritional quality depends on three main component classes: proteins, lipids, carbohydrates. These features, along with the color, determine the commercial value at the beginning of wheat distribution/processing. In fact, protein content, in terms of both quantity and quality (in terms of the gliadin and glutenin content), influences the probability of being processed and the characteristics of the end-product food. As an example, high protein content leads to increased water absorbance and nutritional

value, increasing also the shelf-life of the products. Grain storage proteins include gliadin, glutenin, albumin, and globulin, for which accurate, economical, and rapid assessment could be beneficial for researchers and the industry.

Infrared measurements performed in the spectral region around  $200\text{ cm}^{-1}$  are commonly used for determining the protein content of wheat flour and can be considered a standard industrial method. On the other hand, it has been shown that quantification of protein content using NIR instruments achieves an error in predicting protein content of 0.05% protein on dry matter (d.m.), comparable to benchtop IR instruments (Ye et al. 2018). This result was obtained by developing a Partial Least Square (PLS) model and data from benchtop NIR.

Recently, protein evaluation in intact seeds (barley) using NIR hyperspectral data (Caporaso et al. 2018) and a Convolutional Neural Network (CNN) has reached an error of 0.08% d.m (Singh et al. 2023).

Several studies based on near-infrared spectroscopy investigated the ripening of fruits in recent years. NIR spectroscopy has been used to determine Total Soluble Solids (TSS) concentration (TSS in %) in apples (Nturambirwe et al. 2019; Pourdarbani et al. 2020; Y. Zhang et al. 2019b) mandarins (McGlone et al. 2003), peaches (Kawano et al. 1992), mangoes (Saranwong et al. 2001), table grapes (Donis-González et al. 2020b), and wine grapes (Benelli et al. 2020). To the best of our knowledge, the pomegranate intact fruit has never been tested before with a portable miniaturized NIR device for the determination of its ripeness, and pomegranate fruits have received less attention compared to more common fruits such as apple or citrus (Opara et al. 2022).

Pomegranate (*Punica granatum* L.) is a member of the *Punicaceae* family, and it was one of the earliest fruit crops to be domesticated. It is native to the arid and semi-arid regions of Iran (old Persia) and northern India, and is nowadays cultivated throughout the world from the Mediterranean basin to subtropical and tropical areas of Asia and America.

The pomegranate has a natural appeal related both to its long history of domestication, cultivation, and symbolism (which goes back to the origin of agriculture) and to the health effects attributed to the flower, fruit and other parts of the tree (Sarkhosh et al. 2020). In recent years, several new plantings have been established mainly with the cultivar Wonderful for juice production. To harvest the fruit at the correct ripening time, a maturation index widely used is the total soluble content also known as Total Soluble Solids (TSS) expressed as % or °Brix. The decision on harvesting time is critical for obtaining the best fruit quality either for fresh consumption or for juice production or other processing activities.

A method such as the one proposed in this paper, which combines NIR spectroscopy and machine learning approaches, could therefore be of considerable interest either for decision making in the agro-industrial sector or to limit the food waste along the farm-to-fork chain.

## Materials and methods

### Food matrices

#### Cereal seeds samples

Two species of wheat were evaluated in this study: namely durum wheat, *Triticum turgidum* var. *durum* (hereinafter referred to as Durum Wheat), and *Tritordeum* (durum wheat × wild barley). They differ in protein content covering a range from 8.1 to 19.88% d.m. calculated as:  $[\text{Tot. GPC}\% / (100 - \text{H}\%)] \times 100$ ; the samples derived from different varieties (see Additional file 1: Table S1) for both species and in total twelve wheat samples (grains) have been assayed; each sample is in technical triplicates. Grain protein content of samples was determined for a 3 g sample of wholemeal flour by NIR spectroscopy using a SpectraAnalyzer device (Basic model, Zeuton) spectral range 1400–2500 nm since this is currently used in the industry as reference method (Du et al. 2022).

#### Pomegranate samples

The pomegranate samples were collected in an orchard located in the Puglia region, Southern Italy, in the countryside of Castellaneta marina (Taranto province). The pomegranate cultivar tested in this trial was the Wonderful, trained to ypsilon trellising and the spacing was 3.3 × 5.7 m.

The trial was conducted over two seasons (2020 and 2021) and the harvested fruits were used to evaluate TSS (Total Soluble Solids), in the two years, whereas titratable acidity (TA), pH, polyphenols (as gallic acid) and vitamin C were analyzed for the 2021 samples only. In 2020, a total of 280 fruits were collected and analyzed from 11 July to 15 October on 7 sampling dates; 210 fruits were used for model calibration and 70 fruits for the external model validation. In 2021, 320 fruits were collected and analyzed from July 21 to December 21 (last sampling in post-harvest) on 8 sampling dates; 240 fruits were used for model calibration and 80 fruits for the external model validation. In both years, the fruits were collected at various development and ripening stages and were chosen to be representative of the conditions observed (color, size, phenological stage) in the orchard at each sampling time. Spectral data for each fruit were acquired under laboratory conditions using the SCiO™ mini (v1.2), followed by destructive analysis for the determination of TSS by using a digital refractometer (model HI 96801, Hanna Instruments, Woonsocket, RI, USA); TA and pH were measured for each fruit in 2021 by using an automated

pH-meter (PH-Burette 24, Crison Instruments, Barcelona, Spain). For each sample acquisition, two (in 2020) or three (in 2021) scans per fruit were performed at different locations around the fruit equator, approximately 180° and 120° apart respectively, at a distance < 0.5 cm from the device as indicated by the manufacturer. The juice for the destructive analysis was collected from each fruit.

### Portable sensors

#### Optical multi-spectral measurements using the 'PhasmaFood' sensor

Multi-sensor spectral data were acquired using a prototype developed in the framework of the PhasmaFood project (Fig. 1). The sensor is equipped with a miniaturized commercial UV–VIS spectrometer (C12880MA, Hamamatsu, Japan), a prototype miniaturized NIR sensor with a spectral range 950–1900 nm (Pügner et al. 2016), and a miniaturized RGB-camera (MU9PC-MH, CMOS, Ximea, Münster, Germany). The UV–VIS spectrometer is used both in fluorescence-emission spectroscopy (FLUO) and a diffuse-reflectance visible-spectroscopy (VIS) modality: a long-pass optical filter with cutoff at 400 nm ensures a block of fluorescence-excitation reflection. The NIR and UV–VIS spectrometers apertures and their respective illumination sources (white LED for visible reflectance OSRAM minitopled LCW MVSG.EC, 365 nm wavelength Nichia NSSU100DT for fluorescence excitation and halogen mini lamp for NIR reflectance) are positioned inside the case, pointing at the interface of the observation window (Fig. 1, bottom right), thereby ensuring that recorded spectra come from the same sample area. The PhasmaFood platform also includes a custom-designed Android application to control the sensor by a smartphone and an online open database: spectral measurements from the different spectrometers are sent to the mobile app and then directly to the cloud database (<https://dashboard.phasmafood.eu/>).

The settings for the individual sensors in terms of illumination power and acquisition times for VIS and FLUO signals were optimized for the current samples. Each sample replicate is acquired three times, and the acquisition routine is organized in three steps: VIS, NIR and FLUO. Visible diffuse-reflectance spectral acquisition consists of 10 consecutive acquisitions followed by 10 dark acquisitions; Near infrared (NIR) reflectance spectra acquisition consists of 255 averages and dark acquisitions; fluorescence emission spectra acquisition consists of 10 consecutive acquisitions followed by 10 dark acquisitions. The whole acquisition routine takes only 2 min.

The PhasmaFood sensor was calibrated by conducting a 99% diffuse reflectance ( $S_{\text{white}}$ ) and dark acquisition ( $S_{\text{dark}}$ ) measurement of a white standard material



**Fig. 1** The Phasmafood multisensor prototype (left side, top and bottom pictures) and SCiO spectrometer (right side) with the respective samples (cereals and pomegranate)

(Spectralex White Diffuser WDF-030-95, Lake Photonics, Uhldingen-Muehlhofen, Germany). Spectral measurements were pre-processed as follows.

The FLUO acquired spectral data were pre-processed by subtracting the dark acquisition sample spectrum from the raw sample spectrum for FLUO measurements ( $S_{FLUOraw}$ )

$$S_{FLUO} = S_{FLUOraw} - S_{dark}$$

VIS and NIR measurements were pre-processed by subtracting the  $S_{dark}$  from the raw sample spectrum and then dividing it by the 99% diffuse reflectance white standard spectrum's difference with the dark acquisition of white standard spectrum.

$$S_{VIS \text{ or } NIR} = \frac{S_{VIS \text{ or } NIRraw} - S_{dark}}{S_{white} - S_{dark}}$$

Cereal seeds were placed in a 60 mm-diameter plastic petri dishes completely filled (10 mm thickness). The PhasmaFood' sensor was put on top of the Petri dish with its window in direct contact with grains allowing 50 mm between the sample surface and the illumination and sensors modules. A custom sample holder was designed to maintain optical alignment and contact between the sample and the instrument and to block environmental light.

#### SCiO sensor

The SCiO™ molecular sensor is a spectrometer produced by Consumer Physics Inc. (Tel-Aviv, Israel) used in horticulture to evaluate the physical and chemical properties of fruit, corn, soybean, wheat and other products. It operates in the range of 740–1070 nm, with a sampling

spectral resolution of 1 nm. It has been extensively described in a recent paper when it was used to test the ripeness parameters of a wine-grape cultivar (Ferrara et al. 2022b). The device can be managed by the proprietary software SCiO™ Lab running on smartphones, tablets, and personal computers with a Bluetooth connection. The device was calibrated before every session for correcting data acquisition using the white tile located inside the sensor cover.

#### Spectral analysis

##### Cereal spectral analysis and performance evaluation

NIR spectra acquired by PhasmaFood sensor were used for the data analysis and sample classification. Spectral data have been processed separately according to the species (Durum Wheat, Triticum) or together (Whole set). Different combinations of data, pre-processing methods, and analysis have been considered and evaluated with performance indicators. NIR spectra with different pre-processing methods are shown in the Additional file 1: Fig. S2).

The NIR cereal spectral-pre-processing step has included three different approaches: no pre-processing (Raw Dataset), Smoothing, and Normalization. In the Raw dataset, spectra have been analyzed without any treatment. Smoothed spectra have been obtained applying Savitzky-Golay filter with a second-order polynomial and an 18-point window to remove the high-frequency noise from the data. Different normalization procedures have been tested on the smoothed dataset: unit-value normalization (UVN) spectra have been obtained by applying  $x_{normalized} = (x - x_{minimum}) / (x_{maximum} - x_{minimum})$ ;

Area normalization (AREANORM) have been obtained by applying  $x_{\text{normalized}} = x / (\text{integral of the spectra})$  and Standard Normal Variate (SNV) by  $x_{\text{normalized}} = (x - \text{mean of the spectra}) / (\text{standard deviation of the spectra})$ . All datasets were used to evaluate the effects of pretreatment methods on the final regression results.

After the pre-processing step, the dataset was split into a training set and a test set, with an 80/20 split for the regression step. Scaling was conducted on the datasets using the training-set mean and standard deviation to transform both the training set and the test set.

Partial Least Square (PLS) and Convolutional Neural Network (CNN) (Bian et al. 2018) methods were then used to build a predictive model of protein content and their performances compared.

The CNN used for this approach is fully described in the Additional file 1: Section 1. The number of PLS latent variables was optimized by plotting the predictive performance of the PLS on a reduced dataset against the number of latent variables.

For the PLS and CNN evaluation, a fivefold random shuffle test-training split was used for cross validation. A random split was chosen instead of the k-fold to reduce the risk of overfitting. Data Analysis was conducted using Python v3.9.12 and Sklearn v1.0.2 (Pedregosa et al. 2011).

After a first run of data analysis and results evaluation, it was clear that the regression methods performed better when samples of different species (Durum Wheat and Triticum) were considered separately.

To identify Durum Wheat spectra from Triticum ones, Principal Component Analysis (PCA) on VIS spectra was then conducted. The PCA and Linear Discriminant Analysis (LDA) were implemented in Python v3.9.12 with Sklearn v1.0.2, the LDA was cross-validated with a 5 split stratified k-fold cross-validation.

The permutation feature importance of the CNN model was calculated using the ELI5 python package, which calculates the permutation importance of a feature by shuffling the values of a feature inside the training set and calculating the decrease of the Mean Absolute Error against the normal MAE calculated without shuffling.

#### ***Pomegranate NIR spectral analysis and performance evaluation***

Data collected by the SCiO™ sensor were analyzed using SCiO™ Lab online interface models, whose predictive capability have already been validated in previous work (Li et al. 2018).

The pre-processing step was empirically designed by testing different combinations of techniques and choosing the one which maximized the performances metrics of the overall model. The combinations of the pre-processing techniques were obtained, taking either the first

or the second derivative of the raw signal or the log-transformed signal. Finally, an SNV step was added. The two best combinations were considered for the prediction model construction.

For PLS in this trial, the number of latent variables was chosen by means of a feature of the SCiO™ software that automatically allows users to select the number of parameters for each model. For the validation step, more detailed information on data analysis can be found in recent papers on table and wine grapes (Ferrara et al. 2022a).

## **Results and discussion**

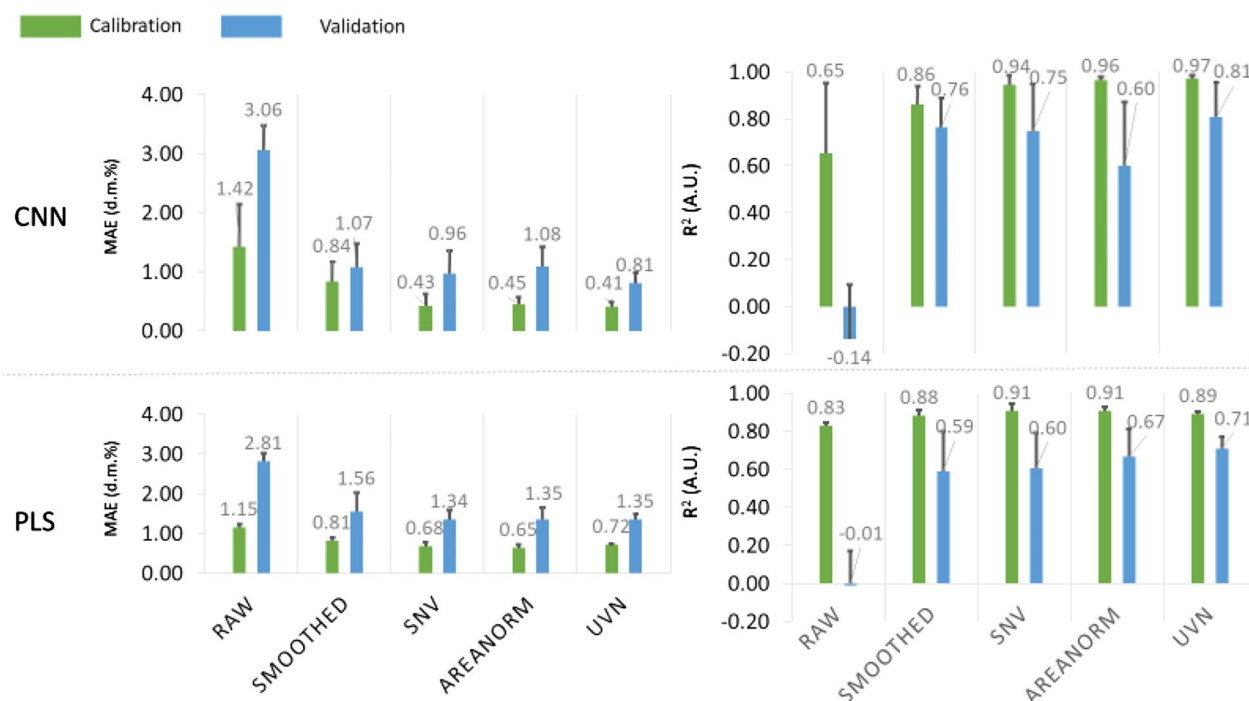
### **Cereal spectral analysis**

The predictive performances of CNN and PLS were evaluated in terms of: Root Mean Square Error (RMSE), Mean Absolute Error (MAE),  $R^2$ , RPD and Ratio of Performance to the Interquartile (that is the interquartile range of the observed values divided by the Root Mean Square Error of the prediction) (Bellon-Maurel et al. 2010). All results are available in the Additional file 1: Table S4 while MAE and  $R^2$  are presented and discussed in this paragraph as representative results of model performance.

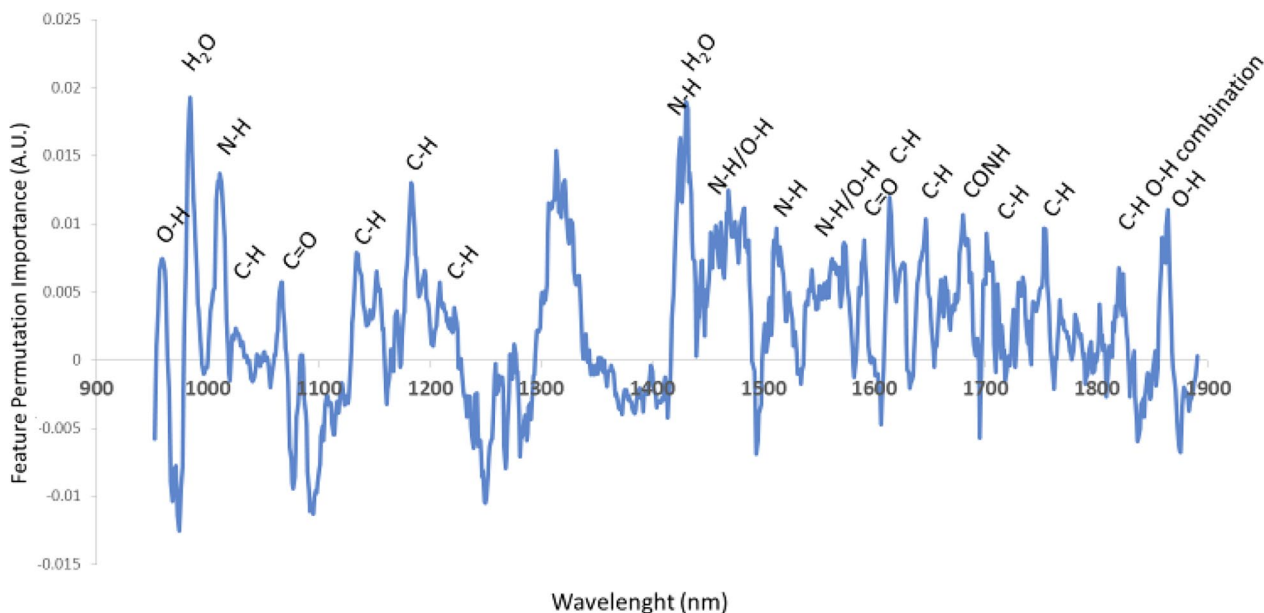
Figure 2 shows the values of  $R^2$  and MAE obtained with CNN and PLS models after different preprocessing treatments to raw data. In this case, samples of both species are considered together, and the results evidence that the CNN model generally outperforms the PLS model.

The importance of data pre-processing is highlighted by the fact that the CNN model cannot accurately predict protein content on the raw validation data and gives a negative  $R^2$ . Regarding the normalization approaches, the unit value normalization improves the predictive performance and produces less overfitting compared to SNV and the area normalization, which have higher errors and a greater difference between Calibration and Validation  $R^2$  that typically indicate overfitting.

Starting from these results, to go deeper into the chemical meaning of the spectral characteristics of the samples, it was necessary to identify the spectral regions of interest (Huan et al. 2021). For this purpose, Feature Permutation Importance analysis (Fig. 3) was used. From this 'black-box' method, we can see the features (wavelengths) that are most important for the CNN model. Notably, the results of Feature Permutation Importance from the CNN model of both species highlighted the same spectral areas as the PLS weights, as shown in the Additional file 1: Fig. S1, but with decreased width of peaks. It is also only a single set of importance values compared to the multiple weights of the PLS. This approach makes easier the identification of peaks related



**Fig. 2** Predictive Performance of % d.m. in cereal samples of CNN and PLS on the dataset when the samples from two species were considered together. Performance in terms of MAE (Mean Absolute Error) and R<sup>2</sup> according to different preprocessing strategies (Raw: no treatment, smoothed: SG smoothing, SNV Standard Normal Variate, Areanorm: area normalization, UVN Unit Value Normalization)



**Fig. 3** Smoothed feature-permutation importance for both species. Standard Normal Variate (SNV), CNN model and peak interpretation

to chemical bonds already identified in NIR literature (Workman and Weyer 2012).

The H<sub>2</sub>O bands of the second and third overtone at 980 nm and 1430 nm have the greatest importance in the model, while the Feature Permutation Importance peaks at 1184 nm, 1510 nm and 1702 nm can be directly attributed to protein content (Workman and Weyer 2012). A detailed interpretation of the peaks is given in the Additional file 1: Table S3), while the combination of bands around 1300 nm is still to be unequivocally identified to our knowledge.

When data analysis is applied to single-species datasets, the predictive performance of CNN and PLS is improved (Fig. 4), and the PLS outperforms the CNN (which fails to predict the protein content in the RAW DW, RAW TD, SMOOTHED TD and UVN TD datasets as indicated by the low R<sup>2</sup>).

Considered separately, the models achieve an error of less than 0.5% d.m. for both species after pre-processing. This result can be considered accurate enough for practical application.

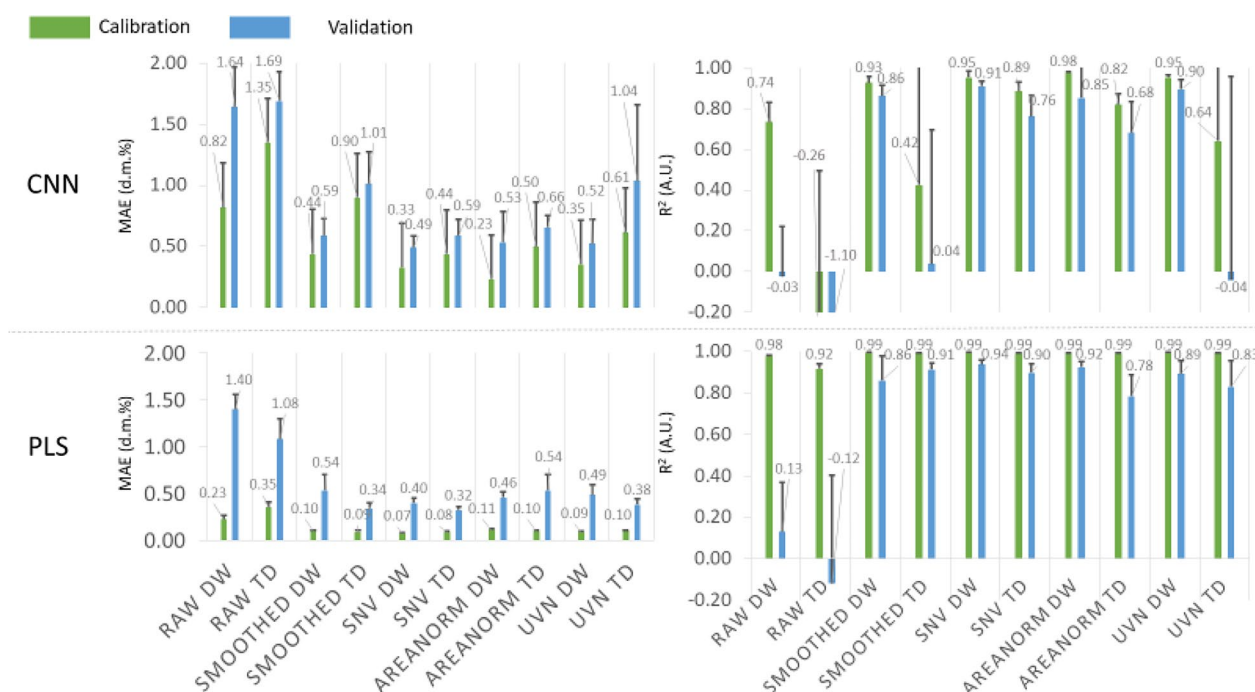
A strategy to take advantage of the better performance of the species-specific model and its wider applicability is to determine the sample species with a discriminant model before predicting the protein content with the appropriate model. In agro-industrial conditions, and at any point of the farm-to-fork value chain, this

methodology can be implemented in a fast and effective way.

To separate spectra belonging to different species using a spectral-analysis method, we considered visible spectra of Durum Wheat (DW) and Triticordeum (TD), acquired during the same acquisition routine for NIR spectra. Applying a PCA to the visible dataset in the score plot of the two principal components PC2 and PC3, which together account for 10.9% of the total variance, it is possible to see a separation between the two types of wheat. This natural separation of the species was observed only in the visible dataset: the mean spectra differ in the region between 525 and 700 nm. To evaluate the possibility of species discrimination, a Linear Discriminant Analysis was trained on the visible dataset. The result is shown in the Additional file 1: Fig. S2, and the accuracy obtained allows the discriminant algorithms to be used.

The use of VIS data in combination with NIR data is an example of a data-fusion approach, where different spectroscopic methods and relative-analysis algorithms related to the same sample are used in combination (sequentially in this case) to improve the prediction accuracy of a model based on spectroscopic data.

The measurable spectral difference in the visible range between Durum Wheat and Triticordeum support the hypothesis that the two cannot be used by the same model for an accurate protein prediction: the considered



**Fig. 4** Predictive Performance of % d.m. in cereal samples of CNN and PLS on the dataset when the two species are considered separately (DW Durum Wheat, TD Triticordeum). Performance is shown in terms of MAE (Mean Absolute Error) and R<sup>2</sup> according to different preprocessing strategies (Raw: no treatment, smoothed: SG smoothing, SNV Standard Normal Variate, Areanorm: area normalization, UVN Unit Value Normalization)

region of NIR spectra doesn't give a direct measure of the protein, but a combination of overtone emissions from a variety of molecular bonds that correlate with the protein content, and the two species have different biology and genes, and consequently proportions in the classes of proteins can differ between the two cereals. It can be hypothesized that the CNN performs better on both species dataset for its ability to isolate the protein signal as it is shown in Fig. 3. Nevertheless, the different behavior of the two species shows that a chemometric analysis should be optimized and should consider every aspect of the nature of the samples and the objective of the classification. Further studies considering both species can lead to improved predictions.

The performances of the CNN and PLS prediction models don't reach the error level of the models trained on flour samples with benchtop instruments (like the one used as reference in this study) but are accurate enough (98% of similarity between the two instruments) to allow the use of the portable instruments for fast non-destructive evaluation of crop quality in the field.

**Pomegranate spectral analysis**

The results of pomegranate fruits can be divided into three parts, in relation to the year of data collection and the different ripening parameter considered:

1. Data collected in 2020 (TSS).
2. Data collected in 2021 (TSS, TA, pH, vitamin C, gallic acid).
3. Data collected in 2020 and 2021 (TSS).

In 2020 (Table 1), the model that has a higher R<sup>2</sup> and RPD and a lower RMSE is the one in which the data have been pre-treated using the logarithm, the second derivative and the SNV, which had an RMSE of 0.98, an R<sup>2</sup> of 0.87 and an RPD of 2.77; the second-best model, however, was obtained using the first derivative and the SNV.

It is possible to compare these values with those obtained by (Ferrara et al. 2022b) for sensor calibration on wine grapes (Primitivo cultivar) trained to Alberello and Guyot in 2019 and 2020.

Table 1 shows an R<sup>2</sup> around 0.81, an indication of a good accuracy in predicting the data of TSS on the validation dataset. In this case, the best model and the best combination of data pre-processing steps is the one that involves the logarithmic transformation, followed by the second derivative and SNV. The RPD of 2.34 and 2.27, being between 2.0 and 2.5, indicates a model capable of making quantitative predictions.

However, the R<sup>2</sup> values for the TSS of this study do not reach the values obtained for wine grape, ≈0.90 (Ferrara et al. 2022b), but pomegranate and grape are two very different fruit species. Only the RMSE, in our case, is better (1.06 and 1.09) than that obtained for wine grape (Ferrara et al. 2022b) which averaged 1.76.

In the case of table grape (Ferrara et al. 2022a), higher values of R<sup>2</sup> and RPD (0.93 and 12.9, respectively) and lower values of RMSE (0.61) were obtained, indicating an excellent predictive model, better than the one obtained in this study for pomegranate.

The difference between grapevine and pomegranate in the accuracy of prediction could be attributed to the thickness of the skin, because in the case of wine- and table grape, this is considerably thinner than that of the pomegranate (which instead has a considerably thicker skin and associated inner tissues for the Wonderful cultivar).

In 2021, however, various parameters were measured in addition to the TSS, such as TA, pH, vitamin C and gallic acid (used as a measure of polyphenols) and the respective predictive models were built for each parameter with the different combinations of pre-treatment, as reported in Table 2.

In 2021, the parameter with the lowest RMSE value is pH, with an average value of 0.10; the TSS is the parameter with the R<sup>2</sup> closest to 1, with an average value of 0.87, and with the highest average RPD, around 2.82.

In the case of TSS, the best model was obtained using raw spectra, while the second-best model was obtained using a combination of two pre-treatment techniques such as the second derivative and SNV.

The R<sup>2</sup> value for the TSS (0.88, Table 2) was clearly higher than values reported in previous studies carried out on other fruit species such as feijoa, apple and

**Table 1** PLSR accuracy on pomegranate TSS by means of the best combinations of pre-processing techniques in the season 2020

Parameter	STD. DEV	Pre-processing combination	RMSE	R <sup>2</sup>	RPD	LVs
TSS 2020						
Calibration	2.70	Log, 2nd Der, SNV	0.98	0.87	2.77	2
		1st Der, SNV	1.00	0.86	2.72	3
Validation	2.48	Log, 2nd Der, SNV	1.06	0.81	2.34	2
		1st Der, SNV	1.09	0.80	2.27	3



**Table 2** PLSR accuracy on pomegranate TSS, TA, pH, vitamin C and gallic acid by means of the best combinations of pre-processing techniques in the season 2021 and 2020–2021 (TSS only)

Parameter	DEV. ST	Pre-processing combination	RMSE	R <sup>2</sup>	RPD	LVs
Calibration						
TSS 2021	3.06	Raw spectra	1.05	0.88	2.93	8
		2nd Der, SNV	1.07	0.88	2.87	5
TA	7.23	Log, 2nd Der, SNV	4.25	0.65	1.70	3
		1st Der, SNV	4.23	0.66	1.71	4
pH	0.24	1st Der, SNV	0.10	0.83	2.41	5
		2nd Der, SNV	0.10	0.84	2.49	5
Vitamin C	1.59	Raw spectra	1.33	0.30	1.20	7
		1st Der, SNV	1.29	0.34	1.23	4
Gallic acid	1.43	Log, 2nd Der, SNV	1.34	0.12	1.07	1
		2nd Der, SNV	1.34	0.12	1.07	1
TSS 20–21	2.94	Log, 2nd Der, SNV	1.07	0.87	2.75	6
		2nd Der, SNV	1.08	0.87	2.73	6
Validation						
TSS 2021	2.93	Raw spectra	1.04	0.87	2.82	8
		2nd Der, SNV	1.28	0.80	2.29	5
TA	7.04	1st Der, SNV	5.43	0.40	1.30	3
		Log, 2nd Der, SNV	5.19	0.46	1.36	4
pH	0.21	2nd Der, SNV	0.15	0.49	1.40	5
		1st Der, SNV	0.15	0.47	1.40	5
Vitamin C	3.15	1st Der, SNV	2.19	0.36	1.44	7
		Raw spectra	2.27	0.32	1.39	4
Gallic acid	3.04	Log, 2nd Der, SNV	3.04	0.06	1.00	1
		2nd Der, SNV	3.04	0.06	1.00	1
TSS 20–21	2.78	Log, 2nd Der, SNV	1.29	0.78	2.16	6
		2nd Der, SNV	1.25	0.80	2.23	6

kiwifruit (Li et al. 2018) or on table grapes (Donis-González et al. 2020a). In other studies, using NIR sensors working in different wavelength ranges (780–1700 nm and 680–1000 nm), the correlation coefficients were 0.60 and 0.68 respectively (Nicolai et al. 2008; Travers et al. 2014).

Higher R<sup>2</sup> values (> 0.9), on the other hand, have been reported for kiwifruits using a spectral range from 800 to 1100 nm (McGlone and Kawano 1998) and for grapes using a spectral range from 800 to 2500 nm (Jarén et al. 2001), but with lab devices.

With regards of the RMSE, the value of 2021 (1.09 on average) was higher than values reported in other experiments, such as for Royal Gala apple (0.72) reported in (McGlone et al. 2002), for pear (0.44) reported in (Nicolai et al. 2008) or for Cotton Candy table grape in 2019 (0.68) and in 2020 (0.59) as detailed in (Ferrara et al. 2022a). Values of RPD of 3.90 and 7.42 were obtained for Cotton Candy table grape in 2019 and 2020, respectively (Ferrara et al. 2022a), which resulted higher than the one obtained for pomegranate in 2021 (2.93).

In the case of pH, both the best models were obtained using the first or second derivative and SNV, with an R<sup>2</sup> equal to 0.83, an extremely low RMSE (0.1) and a fairly high RPD (2.49). The RMSE obtained was very similar to that obtained for Cotton Candy in 2019 in the work of (Ferrara et al. 2022a), while R<sup>2</sup> and RPD resulted much lower.

In the case of titratable acidity (TA), all the models presented an R<sup>2</sup> that was generally low (around 0.64), while in the case of vitamin C and gallic acid the R<sup>2</sup> values observed were extremely low (around 0.27 and 0.08 for vitamin C and gallic acid, respectively) suggesting that the models were not accurate enough.

However, most of the best models seen in Table 2 were obtained by pre-treating the data first with the derivative (first or second) and then with SNV, as also reported by (Donis-González et al. 2020a) for several seedless table grape cultivars.

When testing the models on the validation set, the only parameter with a high R<sup>2</sup> is TSS, as also seen in

2020, with an average of 0.84 between the two best predictive models; the RMSE is around 1.16 on average.

The poor estimation of acidity and pH, and of vitamin C and gallic acid, could be associated with the high thickness of the skin which prevents the penetration of radiation up to the arils, thus creating considerable interference in the determination of such molecules. Moreover, the presence of polyphenols (organic acids in the mesocarp beneath the skin) can be another factor affecting the accuracy of the NIR measurements with respect to soluble solids. Therefore, the only parameter for which a good predictive model was obtained resulted from TSS.

The depth of penetration of NIR applied on fruit depends on the wavelength (Lammertyn et al. 2000), and is about 2–3 mm in the range of 900–1900 nm but can also reach 4 mm at a shorter wavelength (about 700–900 nm). Consequently, knowing the wavelength associated with our SCiO sensor, we can assume that the depth reached in our study is approximately 3–4 mm, given the wavelength of 740–1070 nm. Thus, the spectra data obtained are probably related mainly to the properties of the inner part of the fruit skin and partly of the arils, and this could have greatly influenced the prediction accuracy and the model performance for the different molecules, in particular the ones at low concentration (vitamin C, polyphenols).

Therefore, the correlation between spectra data obtained with the sensor and the data obtained by destructive analysis have given rise to predictive models with a medium–high accuracy in the case of TSS, a medium–low accuracy in the case of TA and pH and a very low accuracy in the case of vitamin C and total polyphenols (as gallic acid). The latter are not usable for practical application considering the low accuracy.

The best combinations among the pre-processing techniques always included the logarithmic transformation (Log), first or second derivative (1st Der or 2nd Der) and SNV (Standard Normal Variate). The major peaks, as can be seen from Additional file 1: Fig. S8, are present at a wavelength of 950–1000 nm, corresponding to the main functional groups such as OH, NH<sub>2</sub>, H<sub>2</sub>O, ROH and ArOH (with OH bond on the aromatic) (Cen and He 2007; Johnson et al. 2020).

These groups are related to major constituents of juices such as sugars, vitamins, polyphenols and other biologically active compounds. TSS mainly represents sugars, but also, to a less extent, other organic molecules (vitamins, amino acids, hormones, etc.) containing C–C, C–O, O–H and C–H bonds, and a NIR spectroscopy can be used for a non-destructive measurement of sugars (Johnson et al. 2020; Ma et al., 2019).

The data of TSS collected both in 2020 and in 2021 were used to create a two-year model. For the creation of

these models, two scans per fruit were performed using the SCiO Lab portal, in such a way as to take into consideration both the 2020 and 2021 samples.

The model with the higher R<sup>2</sup> and RPD (0.87 and 2.75, respectively) and a lower RMSE (1.07) was the one in which the data were processed using the logarithmic transformation, second derivative and SNV.

These values, not as high as in other studies previously reported, could be due to the different growing conditions that occurred in 2020 and 2021, thus leading to substantial differences in the creation of effective predictive models.

A study carried out for 4–5 years, in different growing and pedoclimatic conditions, could help in the construction of more accurate predictive models, with a high R<sup>2</sup> and RPD and a low RMSE. Additional file 1 shows PLS components weights, List of Samples, NIR and VIS Spectra, CNN architecture, peak identification, PCA score plot for wheat samples, classification performances.

## Conclusions

In this paper, potential advantages of using NIR portable devices coupled to machine-learning analysis methods for the qualitative evaluation of intact raw food products have been investigated, together with specific limitations.

In particular, the total protein content in wheat seeds has been determined for *Durum Wheat* and *Triticum aestivum* with a minimum validation error of 0.4% and 0.32% d.m. respectively, and TSS in intact pomegranate fruits has been determined with a RMSE of 1.25.

The application of portable sensors using NIR spectroscopy has the potential to be a valid and rapid alternative to destructive laboratory techniques for the assessment of protein content in intact wheat seeds and ripeness grade in intact pomegranates by evaluating TSS only. It must be always taken into account that NIR is sensitive to the overtones of molecular bond vibrations, that they overlap and are non-specific to a single analyte, so chemometric models trained on this type of data can be case specific and may fail (or significantly reduce regression/classification performance) when applied to different families of samples, as observed in the case of the two wheat species. In these cases, the use of data-fusion approaches, and, in particular, the combination of different spectroscopic data (in our work visible reflectance data beside NIR spectra) can lead to improved results and widen the range of potential applications and samples, as the collection of information from different techniques allows more complex scenarios to be considered in terms of variable parameters.

In the case of pomegranate, not all the parameters tested were successfully predicted by the models, while the same models proved very accurate when applied to

different types of fruits (e.g., grapes) in previous studies. In this case, the limitation of the approach is due to the physically-limited penetration depth of NIR radiation inside pomegranate skin: TSS signal is detectable from the outer layers of pomegranate skin, whereas other molecules at low concentration are not accurately detected. Also in this case, the set of parameters available for accurate detection must be chosen considering the specificity of the samples.

For the possibility of rapid detection, portable spectroscopic devices could be used either in the field by agronomic consultants to quickly assess the ripeness of pomegranate fruits thus reducing costs, personnel labour and use of chemicals or by technicians/farmers on arrival of fruits at the farm to make a first selection of fruits based on ripeness for successive processing. Applications would be possible also in stores by retailers for the on-site monitoring of the ripening status of the fruits, always considering that near infrared light penetrates only a few millimetres so only properties that influence the surface can be measured.

A further miniaturization of devices and/or evolution of the whole platforms in the next future will probably allow the use of this technology for wider application in food safety, authenticity and quality along all the farm-to-fork value chain. The possibility to implement complex chemometric models in cloud-based automatic classification procedures will also allow non-expert users to receive data and inform decision-making in a fast and reliable way.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43170-024-00244-z>.

**Additional file 1.** PLS components weights, List of Samples, NIR and VIS Spectra, CNN architecture, peak identification, PCA score plot for wheat samples, classification performances.

### Acknowledgements

The authors wish to thank Alessandro Giammarino for help in data acquisition, and all PhasmaFOOD team for the great job in designing and producing multisensor prototype.

C. Ricci, and A. Giammarino, supported by Spectrafood project (POR FESR LAZIO 2014-2020, Gruppi di Ricerca 2020, project N. A0375-2020-36643). The work has also been supported by EU-funded Agritech National Centre for Technology in Agriculture (in the framework of Next Generation EU initiative), Spoke 7 and 9.

### Author contributions

C.R., investigation, analysis, original draft preparation; A.Gad., resources, analysis, original draft preparation, writing—review and editing; A.Ger., conceptualization, investigation, original draft preparation, writing—review and editing; A.Did., investigation, analysis; G.F., resources, analysis, original draft preparation, writing—review and editing; F.R.B., conceptualization, investigation, original draft preparation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

### Funding

This work has also been supported by grants from project “Recupero del germoplasma frutticolo pugliese 2.2. – Fruttiferi Minori e Agrumi” con acronimo “REGFRUP 2.2”, PSR Puglia 2014–2022. Misura 10 - Pagamenti agro-climatico- ambientali. Sottomisura 10.2 - Sostegno per la conservazione, l'uso e lo sviluppo sostenibili delle risorse genetiche in agricoltura. Operazione 10.2.1 - Progetti per la conservazione e valorizzazione delle risorse genetiche in agricoltura.

### Availability of data and materials

On request.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 5 January 2024 Accepted: 4 April 2024

Published online: 14 April 2024

### References

- Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, Roger JM, McBratney A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC - Trends Anal Chem.* 2010;29(9):1073–81. <https://doi.org/10.1016/j.trac.2010.05.006>.
- Benelli A, Cevoli C, Fabbri A. In-field Vis/NIR hyperspectral imaging to measure soluble solids content of wine grape berries during ripening. 2020 IEEE International Workshop on Metrology for Agriculture and Forestry, MetroAgriFor 2020 - Proceedings, 99–103. 2020. <https://doi.org/10.1109/MetroAgriFor50201.2020.9277621>
- Bertani FR, Businaro L, Gambacorta L, Mencattini A, Brenda D, Di Giuseppe D, De Ninno A, Solfrizzo M, Martinelli E, Gerardino A. Optical detection of aflatoxins B in grained almonds using fluorescence spectroscopy and machine learning algorithms. *Food Control.* 2020;112:107073. <https://doi.org/10.1016/j.foodcont.2019.107073>
- Bertani FR, Mencattini A, Gambacorta L, De Ninno A, Businaro L, Solfrizzo M, Gerardino A, Martinelli E. Aflatoxins detection in almonds via fluorescence imaging and deep neural network approach. *J Food Compos Anal* 2024;125:105850. <https://doi.org/10.1016/j.jfca.2023.105850>
- Bian X, Diwu P, Zhang C, Lin L, Chen G, Tan X, Guo Y, Cheng B. Robust boosting neural networks with random weights for multivariate calibration of complex samples. *Anal Chim Acta.* 2018;1009:20–6. <https://doi.org/10.1016/j.aca.2018.01.013>.
- Caporaso N, Whitworth MB, Fisk ID. Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chem.* 2018;240:32–42. <https://doi.org/10.1016/j.foodchem.2017.07.048>.
- Cen H, He Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci Technol.* 2007;18(2):72–83. <https://doi.org/10.1016/j.tifs.2006.09.003>.
- Donis-González IR, Valero C, Momin MA, Kaur A, Slaughter DC. Performance evaluation of two commercially available portable spectrometers to non-invasively determine table grape and peach quality attributes. *Agronomy.* 2020;10(1):1–16. <https://doi.org/10.3390/agronomy10010148>.
- Donis-González IR, Valero C, Momin MA, Kaur A, Slaughter DC. Performance evaluation of two commercially available portable spectrometers to non-invasively determine table grape and peach quality attributes. *Agronomy.* 2020;10(1):148. <https://doi.org/10.3390/AGRONOMY10010148>.
- Du Z, Tian W, Tilley M, Wang D, Zhang G, Li Y. Quantitative assessment of wheat quality using near-infrared spectroscopy: a comprehensive review.

- Compr Rev Food Sci Food Saf. 2022;21:2956–3009. <https://doi.org/10.1111/1541-4337.12958>.
- Ferrara G, Marcotuli V, Didonna A, Stellacci AM, Palasciano M, Mazzeo A. Ripeness prediction in table grape cultivars by using a portable NIR device. *Horticulturae*. 2022;8(7):613. <https://doi.org/10.3390/HORTICULTURAE8070613>.
- Ferrara G, Melle A, Marcotuli V, Botturi D, Fawole OA, Mazzeo A. The prediction of ripening parameters in Primitivo wine grape cultivar using a portable NIR device. *J Food Compos Anal*. 2022;114. <https://doi.org/10.1016/j.jfca.2022.104836>.
- Huan K, Chen X, Song X, Dong W. Variable selection in near-infrared spectra: application to quantitative non-destructive determination of protein content in wheat. *Infrared Phys Technol*. 2021;119(September). <https://doi.org/10.1016/j.infrared.2021.103937>.
- Jarén C, Ortuno JC, Arazuri S, et al. Sugar determination in grapes using NIR technology. *Int J Infrared Millim Waves*. 2001;22:1521–1530. <https://doi.org/10.1023/A:1015046908814>.
- Johnson AS, Amuah EB, Brahms C, Wall S. Measurement of 10 fs pulses across the entire visible to near-infrared spectral range. *Sci Rep*. 2020;10(1):1–7. <https://doi.org/10.1038/s41598-020-61620-z>.
- Kawano S, Watanabe H, Iwamoto M. Determination of sugar content in intact peaches by near infrared spectroscopy with fiber optics in intercanopy mode. *Engei Gakkai Zasshi*. 1992;61(2):445–51. <https://doi.org/10.2503/jjshs.61.445>.
- Lammertyn J, Peirs A, De Baerdemaeker J, Nicolai B. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Biol Technol*. 2000;18(2):121–32. [https://doi.org/10.1016/S0925-5214\(99\)00071-X](https://doi.org/10.1016/S0925-5214(99)00071-X).
- Li M, Qian Z, Shi B, Medicott J, East A. Evaluating the performance of a consumer scale SciO™ molecular sensor to predict quality of horticultural products. *Postharvest Biol Technol*. 2018;145(March):183–92. <https://doi.org/10.1016/j.postharvbio.2018.07.009>.
- Ma YB, Babu KS, Amamcharla JK. Prediction of total protein and intact casein in cheddar cheese using a lowcost handheld short-wave near-infrared spectrometer. *LWT* 2019;109:319–326. <https://doi.org/10.1016/j.lwt.2019.04.039>.
- McGlone VA, Kawano S. Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biol Technol*. 1998;13(2):131–41. [https://doi.org/10.1016/S0925-5214\(98\)00007-6](https://doi.org/10.1016/S0925-5214(98)00007-6).
- McGlone VA, Jordan RB, Martinsen PJ. Vis/NIR estimation at harvest of pre- and post-storage quality indices for "Royal Gala" apple. *Postharvest Biol Technol*. 2002;25(2):135–44. [https://doi.org/10.1016/S0925-5214\(01\)00180-6](https://doi.org/10.1016/S0925-5214(01)00180-6).
- McGlone VA, Fraser DG, Jordan RB, Künnemeyer R. Internal quality assessment of mandarin fruit by vis/NIR spectroscopy. *J Infrared Spectrosc*. 2003;11(5):323–32. <https://doi.org/10.1255/jnirs.383>.
- McGlone VA, Fraser DG, Jordan RB, Künnemeyer R. Internal quality assessment of Mandarin fruit by vis/NIR spectroscopy. *J near Infrared Spectrosc*. 2017;11(5):323–32. <https://doi.org/10.1255/JNIRS.383>.
- Müller-Maatsch J, Alewijn M, Wijten M, Weesepeel Y. Detecting fraudulent additions in skimmed milk powder using a portable, hyphenated, optical multi-sensor approach in combination with one-class classification. *Food Control*. 2021;121. <https://doi.org/10.1016/j.foodcont.2020.107744>.
- Nallan Chakravartula SS, Moscetti R, Bedini G, Nardella M, Massantini R. Use of convolutional neural network (CNN) combined with FT-NIR spectroscopy to predict food adulteration: a case study on coffee. *Food Control*. 2022;135. <https://doi.org/10.1016/j.foodcont.2022.108816>.
- Nicolai BM, Verlinden BE, Desmet M, Saevens S, Saeys W, Theron K, Cubeddu R, Pifferi A, Torricelli A. Time-resolved and continuous wave NIR reflectance spectroscopy to predict soluble solids content and firmness of pear. *Postharvest Biol Technol*. 2008;47(1):68–74. <https://doi.org/10.1016/j.postharvbio.2007.06.001>.
- Nturambirwe JFI, Nieuwoudt HH, Perold WJ, Opara UL. Non-destructive measurement of internal quality of apple fruit by a contactless NIR spectrometer with genetic algorithm model optimization. *Sci Afr*. 2019;3. <https://doi.org/10.1016/j.sciaf.2019.e00051>.
- Opara UL, Arendse E. Near-infrared spectroscopy for pomegranate quality measurement and prediction. In *nondestructive quality assessment techniques for fresh fruits and vegetables*. Singapore: Springer Nature Singapore; 2022, pp. 211–232.
- Osborne BG. Near-infrared spectroscopy in food analysis. *Encyclop Anal Chem*. 2006. <https://doi.org/10.1002/9780470027318a1018>.
- Pedregosa F, Weiss R, Brucher M, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12(85):2825–30.
- Pourdarbani R, Sabzi S, Kalantari D, Arribas JI. Non-destructive visible and short-wave near-infrared spectroscopic data estimation of various physicochemical properties of Fuji apple (*Malus pumila*) fruits at different maturation stages. *Chemom Intell Lab Syst*. 2020;206(August). <https://doi.org/10.1016/j.chemolab.2020.104147>.
- Pügner T, Knobbe J, Grüger H. Near-infrared grating spectrometer for mobile phone applications. *Appl Spectrosc*. 2016;70(5):734–45. <https://doi.org/10.1177/0003702816638277>.
- Saranwong S, Sornsrivichai J, Kawano S. Improvement of PLS calibration for Brix value and dry matter of mango (using information from MLR calibration). *J Near Infrared Spectrosc*. 2001;9(4):287–95. <https://doi.org/10.1255/jnirs.314>.
- Sarkhosh A, Yavari A, Zamani Z. The pomegranate: botany production and uses. Wallingford: CABI Publishing; 2020. p. 600.
- Singh T, Garg NM, Singh V. Near - infrared hyperspectral imaging for determination of protein content in barley samples using convolutional neural network. *J Food Measur Charact*. 2023. <https://doi.org/10.1007/s11694-023-01892-x>.
- Travers S, Bertelsen MG, Kucheryavskiy SV. Predicting apple (cv. Elshof) postharvest dry matter and soluble solids content with near infrared spectroscopy. *J Sci Food Agric*. 2014;94(5):955–62. <https://doi.org/10.1002/jsfa.6343>.
- Workman Jr J, Weyer L. Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy. *Journal of Chemical Information and Modeling* (2nd ed.,). Boca Raton, FL, USA, CRC Press (Taylor & Francis group). 2012. <https://doi.org/10.1017/CBO9781107415324.004>.
- Ye D, Sun L, Zou B, Zhang Q, Tan W, Che W. Non-destructive prediction of protein content in wheat using NIRS. *Spectrochimica Acta - Part a: Molecular and Biomolecular Spectroscopy*. 2018;189:463–72. <https://doi.org/10.1016/j.saa.2017.08.055>.
- Zhang X, Lin T, Xu J, Luo X, Ying Y. DeepSpectra: an end-to-end deep learning approach for quantitative spectral analysis. *Anal Chim Acta*. 2019;1058:48–57. <https://doi.org/10.1016/j.aca.2019.01.002>.
- Zhang Y, Nock JF, Al Shoffe Y, Watkins CB. Non-destructive prediction of soluble solids and dry matter contents in eight apple cultivars using near-infrared spectroscopy. *Postharvest Biol Technol*. 2019;151:111–8. <https://doi.org/10.1016/j.postharvbio.2019.01.009>.
- Zhong J, Wang X. Evaluation technologies for food quality. *Eval Technol Food Qual*. 2019. <https://doi.org/10.1016/c2017-0-01187-4>.
- Zhou L, Tan L, Zhang C, Zhao N, He Y, Qiu Z. A portable NIR-system for mixture powdery food analysis using deep learning. *Lwt*. 2022;153. <https://doi.org/10.1016/j.lwt.2021.112456>.
- Žilić S, Barać M, Pešić M, Dodig D, Ignjatović-Mičić D. Characterization of proteins from grain of different bread and durum wheat genotypes. *Int J Mol Sci*. 2011;12(9):5878–94. <https://doi.org/10.3390/ijms12095878>.
- Jaren C, Ortuno JC, Arazuri S, Arana I, Salvadores MC. Sugar Determination in Grapes Using NIR Technology. *Int J Infrared Millim Waves*. 2001;22(10):1521–1530. <https://doi.org/10.1023/A:1015046908814>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.