



Adulterant estimation in paprika powder using deep learning and chemometrics through near-infrared spectroscopy

Wilson Castro¹ · Jimy Oblitas² · Luis Nuñez¹ · Ives Yoplac³ · Himer Avila-George⁴ · Miguel De-la-Torre⁴ 

Received: 12 September 2023 / Accepted: 12 April 2024 / Published online: 6 May 2024

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract

Spices and other food products have been permanently susceptible to adulteration, affecting safety and acceptability when commercialized. A relevant alternative to detect contaminants in food products is to couple near-infrared spectroscopy (NIR) with chemometrics. Among the most accurate chemometric techniques employed to analyze food products, partial least squares regression (PLSR) combines features from and generalizes principal component analysis (PCA) to create compact and accurate models. Other techniques inspired in the human brain, such as multilayer perceptron, the long short-term memory (LSTM) models, and other approaches based on deep learning, take advantage of the high complexity of weights and neurons to train models based on large amounts of data. In this paper, a methodology is proposed to evaluate chemometric tools to estimate the percentage of adulterants in paprika powder using NIR spectroscopy, and three approaches are proposed and compared showing different performances. According to the methodology, the paprika samples were dried and separated into pericarp, peduncle, and seed cake. The resulting elements were finely milled, sieved, and mixed into 21 different combinations with a different percentage of each. Spectral profiles were used to train PLSR, multilayer perceptron, and regression models based on LSTM networks. The models were compared following a k -fold cross-validation strategy. Results showed that PLSR presented the highest $R^2 = 0.978$ for peduncle adulterant estimation, and the lowest $RMSE = 6.24$. In particular, when seed cake powder was used as an adulterant, the PLSR approach showed the highest $R^2 = 0.981$, and the lowest $RMSE = 5.806$. The RPD values were higher than 2.000 for all models that use the peduncle as an adulterant and only for models bound to the PLSR in the adulterated samples with pressed seed cake. In summary, the best predictions were obtained using PLSR models, providing evidence of the feasibility of using NIR spectra to estimate the percentage of adulterants in paprika powder.

Keywords LSTM · NIR · Paprika pepper · Perceptron · PLSR

1 Introduction

Adulteration is a growing concern that impacts enterprises, producers, consumers, and the economy in general [1]. Estimated losses due to adulteration of products are higher than USD\$250 billion, and in the food industry, the losses exceed USD\$49 billion in adulterated products [2]. Adulteration includes the replacement of the original material with lower-cost products, defective material, or residues of the same or different plants, harmful substances, or synthetic products that do not meet official standards. Food adulteration may be either intentional (*direct*) aiming to obtain a financial profit or unintentional (*indirect*) due to a

defective production process. In any case, food adulteration represents fraud and therefore constitutes an illegal practice [3].

Among the food products susceptible to adulteration, spices and condiments are high-valued products with an international market, and are widely used as flavorings and food and beverage coloring [4]. Regarding the spices market, between 2014 and 2015, the US FDA (Food and Drug Administration) reported more than 20 adulteration cases in condiments and food spices, causing their withdrawal from the market due to latent danger to consumers [5, 6]. The previous studies have identified cassava, corn starch, and wheat as frequent adulterants in powdered condiments that include garlic, ginger, onion, black pepper, cumin, and seasonings in general. As an example, recent

Extended author information available on the last page of the article

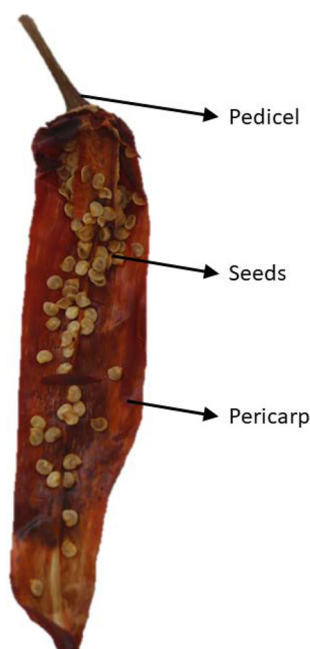


Fig. 1 Paprika chili parts used in the study

studies have identified starch as the main adulterant used in black pepper (up to 50% of dry weight) [7]. Cheaper adulterants are added to obtain an economic benefit by diluting more valuable ingredients, at the expense of consumers' health [5, 7].

Among the most popular spices, paprika (*Capsicum annuum* var. Longum) is a Native American spice that is currently cultivated throughout the world. Originally, the paprika powder was obtained from the pericarp of the fruit to be used as food coloring, seasoning, and flavoring. Currently, oleoresin is also extracted from this fruit to be used as a natural dye in the food, poultry, dairy, feed, canning, bakery, and cosmetic industries, among others [8]. Paprika oleoresin is characterized by its high content of vitamin C and carotenoids, β -carotene, β -cryptoxanthin, capsanthin, capsorubin, violaxanthin, and other carotenoids that are the pigments responsible for their characteristic colors, while capsaicinoids are the spicy compounds present in the fruit [9, 10].

According to Guillen et al. [11], the anatomical parts of the *Capsicum annuum* fruit are the peduncle, pericarp, seed, and placenta. The main part used to obtain the powdered dye should be the pericarp, and it is commonly adulterated with the addition of other parts of the same fruit (e.g., peduncle, seed, and placenta) to reduce production costs [4]. Furthermore, the fraudulent addition of other materials that are not considered food additives has been identified, for example, lead oxide and synthetic dyes [12]. Such illicit actions generate exorbitant profits for those who practice them and represent a significant threat to the

health of consumers [3, 4]. The health hazard that represents an impure food product, encourages the adoption of robust strategies to detect fraudulent adulteration.

Numerous techniques have been developed to detect and estimate adulteration that vary depending on the product, the method employed to measure discriminative properties, and the techniques employed to analyze samples. Table 1 presents some examples of popular adulteration detection methods.

Among the techniques mentioned in Table 1, spectroscopy in general, but especially near-infrared spectroscopy (NIRS), is attracting interest due to the advantages that include low cost, simplicity, non-destructive measurements, and quickness [4, 17]. NIRS measures the molecular vibrations of target products by light quanta absorption, which generates a signature of the spectral profile ("fingerprint") that is reproducible, is distinct for different raw materials, and, in many cases, can be employed to determine the purity or the level of adulteration in food products [2]. Putting it differently, molecular vibrations are related to the conformation, structure, molecular interactions, and chemical bonds of materials, measuring chemical bonds based on overtones and combination bands of specific functional groups [29, 30]. The previous studies report the usefulness of NIRS to ensure effective food supply surveillance and its ability to reduce or detect food forgery or adulteration, mainly in the spectral range between 780 and 2500 nm (4000 and 14 000 cm^{-1}) [2, 19, 20, 31].

On the opposite, NIRS presents some drawbacks, e.g., the wide absorption band, weak absorption peaks, serious multicollinearity, and, in some cases, the geometry of the fruits, presenting distinct reflection artifacts [32]. To address these problems, chemometric techniques are commonly employed to extract information using tools inherited from signal and image processing (e.g., pre-processing before creating a model), the extraction or selection of variables, and, finally, modeling the relationship between the input variables and the properties to be measured [33]. Likewise, it is well known that for a better practical effect, it is necessary to select wavelengths and use pre-processing methods to remove non-informative variables, producing that way simpler and more accurate models [34–37]. Some of the most popular and efficient models in NIRS are principal component regression (PCR) and partial least squares regression (PLSR) [38]; additionally, the nonlinearity of neural networks is commonly advantageous for certain problems [39]. In fact, due to the high dimensionality and spatial complexity of the matrices extracted from food products using NIRS, in certain applications, it is important to use tools to extract nonlinear

Table 1 Common techniques to detect and estimate adulteration in food products

Method	Technique	Sample	Application	References
Physical	Macroscopic analysis	Herbal drugs	Authentication	[13]
	Microscopic analysis	Dietary supplements	Authentication	[14]
	Tandem mass spectrometry	Spices and herbs	Adulteration	[15]
	Spectroscopy and Chemometrics	Spices, herbs, and fruits	Adulteration	[16]
		Goat milk powder	Adulterant	[4, 17–22] [23]
Analytical	Chromatography	Spices and herbs	Authentication	[4]
	Electrophoretic	Herbal medicines	Authentication	[24] [4]
Biotechnological	DNA differentiation	Chinese medicinal stuffs	Identification	[25]
		Spices and herbs	Authentication	[4]
Immunological	ELISA test	<i>Panax</i> species	Identification	[26]
		Chinese medicinal plant	Authentication	[27]
Biosensor	Nanotechnology	Spices and herbal products	Authentication	[28]

information and self-organizing methods commonly provide efficient solutions.

Relevant examples of chemometric models to deal with nonlinearities are the neural network approaches, such as well-known multilayer perceptron, and the recurrent neural networks (RNN) [40]. In particular, among the RNN models, the long short-term memory (LSTM) networks preserve information from the previous states in hidden layers and use it in the prediction and classification processes [41]. Initially developed for speech recognition, LSTM networks have aroused great interest due to their potential for applications in spectrum discrimination [42]. The combination of RNN and spectral profiles has allowed addressing various tasks related to food, such as the prediction of storage time in black tea [43], the quantification of *Clostridium sporogenes* spores in food products [44], and the detection of moisture levels in individual corn seeds [45], among others.

In this paper, a methodology is proposed for the evaluation of chemometric techniques that employ NIRS to estimate the percentage of two common adulterants in Paprika powder, e.g., peduncle or seeds. Additionally, three of the most popular chemometric techniques were compared using the proposed methodology, demonstrating that NIR spectroscopy can be employed to detect adulterants that are part of the ground red peppers used to produce Paprika powder. Finally, the experiments provide evidence of the feasibility of using partial least squares regression, multilayer perceptron, and LSTM networks to estimate the percentage of adulteration in Paprika powder.

2 Materials and methods

2.1 Raw material

A local spice producer provided a sample composed of whole mature (Paprika) ground red peppers. The sample was cleaned, and all fruits with visual defects on the surface were removed. The resulting materials were stored in plastic bags, hermetically sealed and in dark conditions, to reduce the possibility of damage from moisture or light.

2.2 Experimental methodology

The experimental methodology performed in the present study is shown in Fig. 2 and detailed in subsequent subsections. In general, samples were prepared to measure separately the target product (pericarp) and adulterants (pedicel or peduncle, and seeds cake). Then, NIR spectra profiles were extracted, and a standard pre-treatment was applied to measurements. Models were built from such samples, and performance metrics were computed for comparison.

2.3 Sample preparation

Eleven kilograms of Paprika ground peppers were dried in dark conditions, and at 60 °C, for 30 days, until 10% moisture was obtained. The fruits were then divided into pericarp, seeds, and peduncle, removing all the placenta from the previously separated parts.

The pericarp and peduncle were milled, using a hammer mill company SRL at 1450 RPM, and sieved with a 4-mm ASTM sieve. The seeds were passed through the oil

Fig. 2 Experimental procedure to evaluate the paprika adulterant prediction models

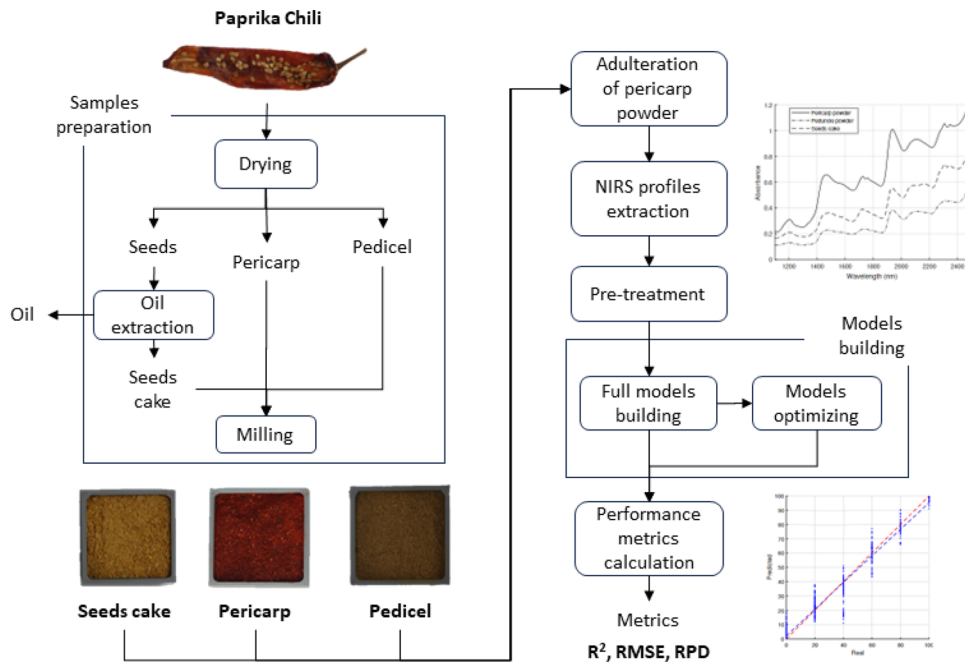


Table 2 Experimental treatment for each sample case

Part	Treatments									
	1	2	3	4	5	6	7	8	9	10
Pericarp	100	80	80	60	60	60	40	40	40	40
Peduncle	0	20	0	40	20	0	60	40	20	0
Seeds cake	0	0	20	0	20	40	0	20	40	60

Part	Treatments										
	11	12	13	14	15	16	17	18	19	20	21
Pericarp	20	20	20	20	20	0	0	0	0	0	0
Peduncle	80	60	40	20	0	100	80	60	40	20	0
Seeds cake	0	20	40	60	80	0	20	40	60	80	100

The 21 treatments are sequentially numbered, and the rest of the quantities represent the percentage of each component (pericarp, peduncle, and seeds cake)

extraction process, the oil was separated, and the seed cake was milled similarly to the pericarp and peduncle. The powder produced for each subproduct was stored separately in hermetically sealed plastic bags and maintained under dark conditions.

2.4 Adulteration of samples

Different treatments are prepared as shown in Table 2, where the target pericarp powder was adulterated with peduncle and seeds cake in different percentages. In both cases, adulterated samples were made using an ultra-turrax model *IKA T25*, at 4500 rpm per 1 min.

2.5 NIRS profiles extraction

The different treatments, consisting of a pericarp, a peduncle, and a seed, were measured in 30 repetitions each, obtaining a total of 630 samples. The spectral profile was determined for each of the 630 measurements.

The measurement of each sample followed the methodology reported by Yoplac et al. [33]. In this study, a Unity Scientific NIR spectrometer (SpectraStar 2500XL, USA) was used, equipped with a tungsten halogen lamp as a light source and an InGaAs detector (Indium–Gallium–

Arsenic) in the range of 1100 and 2500 nm, with a resolution of 1 nm.

Measurements were made in reflectance mode applied directly to mixtures without pre-treatment or manipulation, using a quartz cuvette of 3.5 cm internal diameter and 1.0 cm thick, to which 3.2 g ± 0.3 g of the sample was added.

2.6 Pre-treatment

As reported in the publications by [39, 46], in most cases, extracted spectral profiles contain noise and variability due to capture conditions, and spectral enhancements are required to clean up the profiles, such as spectral smoothing, centering, and normalization.

In this process, the following combinations of standard pre-treatments were applied:

- *Smoothing.* The spectra were smoothed using a second-order Savitzky–Golay filter with eleven frames according to Eq. (1):

$$x' = \frac{1}{N} \sum_{\lambda=1}^n C_{\lambda}(x_{\lambda}), \tag{1}$$

where x' is the smoothed profile; x is the original spectra; C is the coefficient; λ is the wavelength in analysis; and N is an integer number of convolutions.

- *Centering and normalization.* The distribution of samples is centered and normalized according to Eq. (2), to reduce the variation in the baseline due to the dispersion of light.

$$x'_{\lambda} = \frac{x_{\lambda} - \bar{x}}{S_{x,\lambda}}, \tag{2}$$

where x , x' are the original and corrected profiles, respectively; λ is the wavelength to be analyzed; and $S_{x,\lambda}$ is the standard deviation of the profiles at a specific wavelength.

2.7 Models training

The models were trained by implementing functions and routines in the mathematical software MATLAB-2022^a; dividing this stage into the two steps detailed in Sects. 2.7.1 and 2.7.2.

2.7.1 Full models training

Using the profiles corrected in the previous stage, we proceeded to train adulteration–prediction models; for one or two adulterants at a time. The models implemented included the following:

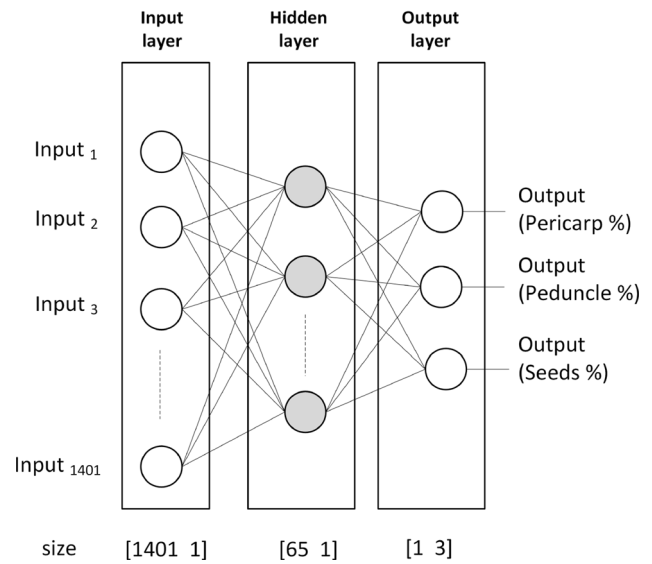


Fig. 3 Architecture of the multilayer perceptron employed in experiments

- *Partial least squares regression (PLSR).* This is one of the most widely used methods to predict food properties by coupling them to hyperspectral images, an example of which includes vibrational spectrometry [39, 47]. PLSR transforms an input matrix X , in our case with dimensions $[m \times n]$, where m is the number of observations, and n is the number of wavelengths. The output vector Y , which contains the percentage of adulterant quality, is obtained by decomposition. Decompose X and Y , by projection, into new directions with the constraint that the decomposition must describe the change of both variables as much as possible. After the decomposition of the variables, a regression step is performed in which the decomposed X and Y are used to calculate a regression model called the full model [48], see Eq. (3).

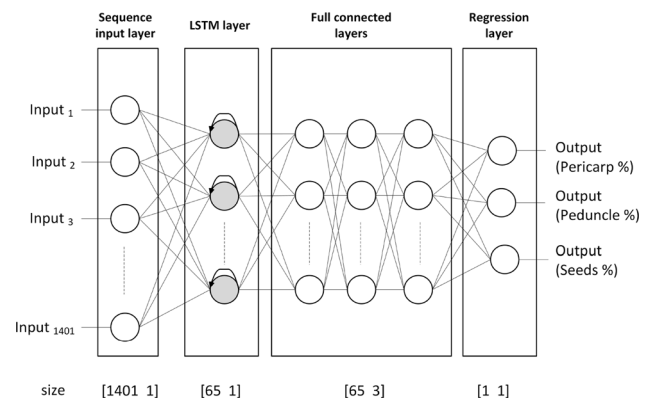


Fig. 4 Structure for the RNN-LSTM regression model

$$Y = \beta.X + e, \quad (3)$$

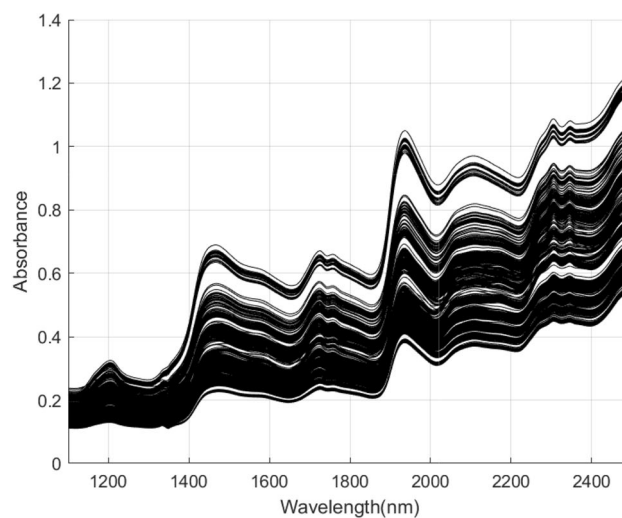
where X and Y are the input and output variables, β is the vector of regression coefficients, and e is the error. For the implementation of this model, the *PLS toolbox* of MATLAB was used.

- **Multilayer perceptron (MLP).** This type of supervised learning network is widely used for prediction and classification, which is generally composed of three layers [39]. The input layer receives the intensity values, which employing a transfer function are distributed to the processing elements (neurons) of the second layer or hidden layer. Commonly, in the second layer, the values that were entered are transformed by a nonlinear sigmoid transfer function, propagating them to the third layer or output layer. The prediction results are obtained at the output layer [49]. The architecture of the multilayer perceptron is depicted in Fig. 3
- **Recurrent neural network (RNN)-based regression.** The regression model based on recurrent networks used long short-term memory (LSTM) networks. Following the pyramidal principle proposed by Vázquez et al. in [39], a model of an input layer with i entering neurons, an LSTM layer with j neurons units, a fully connected layer, and a regression layer, this network structure is illustrated in Fig. 4. The training was carried out using the stochastic gradient descent with momentum (SGDM) algorithm, calculating the gradients in the weights and adjusting them to minimize the loss function, in 600 epochs at a learning rate ($\text{LearnRate} = 0.005$). The training was carried out 30 times in a k -fold ($K = 5$) cross-validation strategy, and its results were stored on each occasion to later proceed to the calculation of metrics.

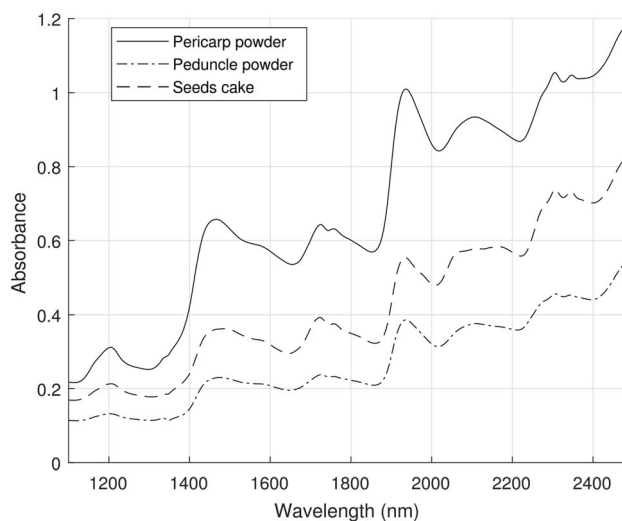
2.7.2 Model optimization

According to Blanco and Villarroya [29], the analytical information contained in the wide and often overlapping bands in the NIR spectrum is hardly selective. For this reason, it is important to choose relevant bands in actual chemometric applications. In this sense, relevant variables were determined.

This step starts with selecting the relevant variables; although there are a wide number of methods for this experiment, the β coefficient method was used. β coefficient is based on the ability of the variables to contribute to the PLSR regression model, defined by their coefficients β and following the strategy applied by [48]. Consequently, the result of this stage is a new spectral profile, named trimmed profiles, which contains only the intensity values for the relevant variables.



(a)



(b)

Fig. 5 Spectral profiles for **a** the whole sample including repetition and **b** average sample profiles for each paprika part (e.g., peduncle, pericarp, and seed cake)

Finally, the models were optimized according to the work of [33, 39], and new models were [48] trained with the optimized trimmed profiles.

2.8 Models comparison

The different models were tested using a k -fold cross-validation strategy with $k = 5$. Then, in the same way as [3, 33, 39], the root-mean-square error ($RMSE_{cv}$), the coefficient of determination (R^2_{cv}), and the ratio of performance to deviation (RPD_{cv}) were computed. These statistical metrics are defined in Eqs. (4)–(6).

$$RMSE_{cv} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \tag{4}$$

$$R^2_{cv} = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}, \tag{5}$$

$$RPD_{cv} = \frac{SD}{RMSE_{cv}}, \tag{6}$$

where \hat{y}_i and y_i are the percentages of adulteration values of the i th sample for prediction and reference, respectively; n is the number of samples; and SD is the standard deviation of Y . The sub-index $_{cv}$ makes reference that the statistical measure was computed following the cross-validation strategy.

3 Results and discussion

3.1 NIR spectra profile

Figure 5a shows all spectra profiles collected from the different samples, including distinct treatments and repetitions; a direct relation between wavelength and absorbance was observed. In Fig. 5b, the differences between the mean spectral profiles of the materials used are observed, showing that the peaks of the three spectral profiles appear at similar locations. The difference between the average profiles is the average level of absorbance between the pericarp, peduncle, and the seed cake. The pericarp spectral profiles present higher absorbance levels, followed by the peduncle, and the seed cake profile with the lowest absorbance. Similarly, nine picks are common in all parts around wavebands 1205, 1460, 1725, 1761, 1930, 2100, 2303, 2351, and 2485 nm. Among these peaks, the first five are related to the absorbance bands in the overtone region, and the remaining four are in the combining region.

The absorption peaks at 1460 and 1930 nm are related to –OH stretch and –OH stretch/deformation wavebands combination, respectively, those mainly due to the presence of water in the samples. The peak at 2100 nm is observed to be within the waveband of –NH deformation, associated with the presence of proteins and/or peptides; this same peak at 2100 nm is within the range of C–O and O–H stretching combination, related to the presence of carbohydrates. The peaks 1205, 1725, and 1761 nm are observed to be within the ranges of –CH₂ and –CH₃ stretch, related to the presence of lipids. Finally, the peaks 2303, 2352, and 2485 nm correspond to the wavebands of methylene and –CH stretch, related to the presence of lipids [50].

The absorption bands of capsicum composition can be analyzed and identified by their spectral behavior; when

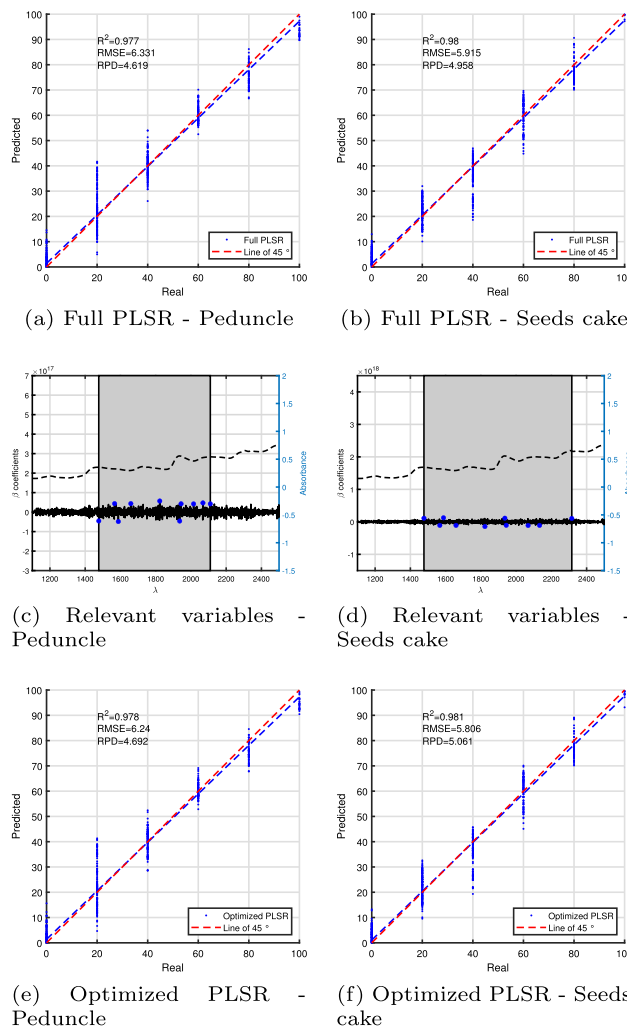


Fig. 6 Evaluation of the PLSR-based models for adulterant prediction

mixed with other components, such as water, carbohydrates, proteins, and lipid content, the characteristics may change. This observation is supported by research papers, such as [3], which evaluate the adulteration of spices using NIR to detect spectra that can increase or decrease depending on the variation in composition. Research articles such as [11] evaluate the pericarp and the non-edible portion (seeds, placenta, and interocular septum) of two capsicum varieties, finding differences in the amounts of capsaicin and dihydrocapsaicin. These differences may explain the variations in spectra in the NIR range observed in our study. This, coupled with the highly sensitive nature of NIR spectroscopy, gives it the potential to differentiate mixtures, but also the potential to detect differences between varieties. Studies like [9], that use HPLC, show the variations in the carotenoid composition in capsicum varieties that can be determined using NIRS.

Whereas the spectral differences can be attributed to the composition of the mix, there are no specific functional

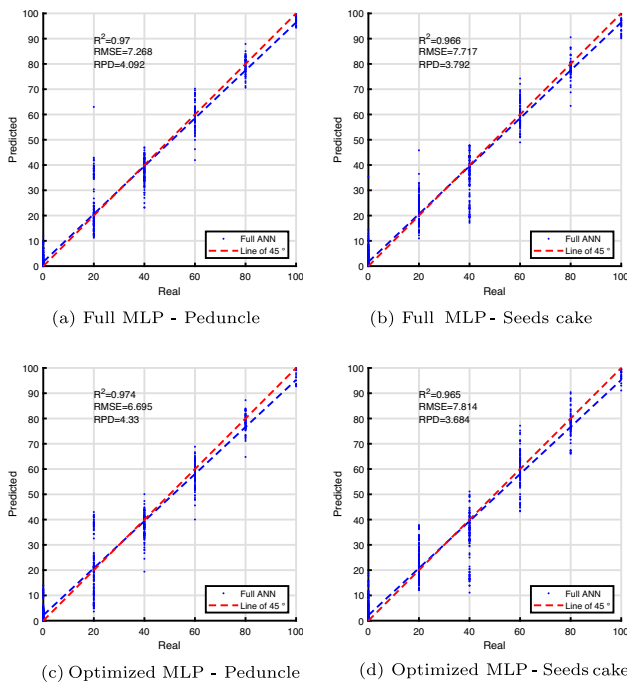


Fig. 7 Multilayer perceptron models for adulterants prediction

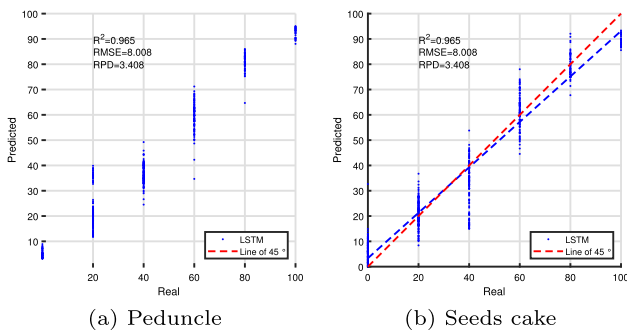


Fig. 8 LSTM models for adulterants prediction

groups that qualitatively differentiate the capsicum samples without an appropriate statistical multivariate analysis.

3.2 Models building

3.3 Partial least square regressions

Figure 6 shows the results of the analysis after the application of the PLSR model to predict the different percentages of combinations of adulterated pericarp seed with peduncle and seed cake. In Fig. 6a and b, the regression of the adulterated pericarp powder with different percentages of the peduncle using all the wavelengths of the NIR spectrometer. On the other hand, the more relevant wavelengths used to generate the optimized PLSR model are shown in Fig. 6c and d. Finally, the real against the

estimated levels of adulteration are shown in Fig. 6e and f, when different adulterants are added to the mixture, but optimizing PLSR.

Regarding the most relevant wavelengths, those in the range between 1600 and 2000 nm stand out, being the most prominent absorption bands at 1725 nm and 1761 nm, which are related to the C=O functional group. Similar wavelengths have already been documented and linked to capsaicin and other capsaicinoids [10]. Partitioning the full spectrum to reduce random noise and computational complexity while extracting the available information to enhance the model’s capacity is a recommended practice for achieving better prediction models [34].

3.4 Multilayer perceptron models

On the other hand, Fig. 7 shows the results of multilayer perceptron models, both the complete and optimized models, to predict the adulteration with the adulterants in pericarp powder; Fig. 7a–d for peduncle adulteration, and Fig. 7b–d when seed cake is used.

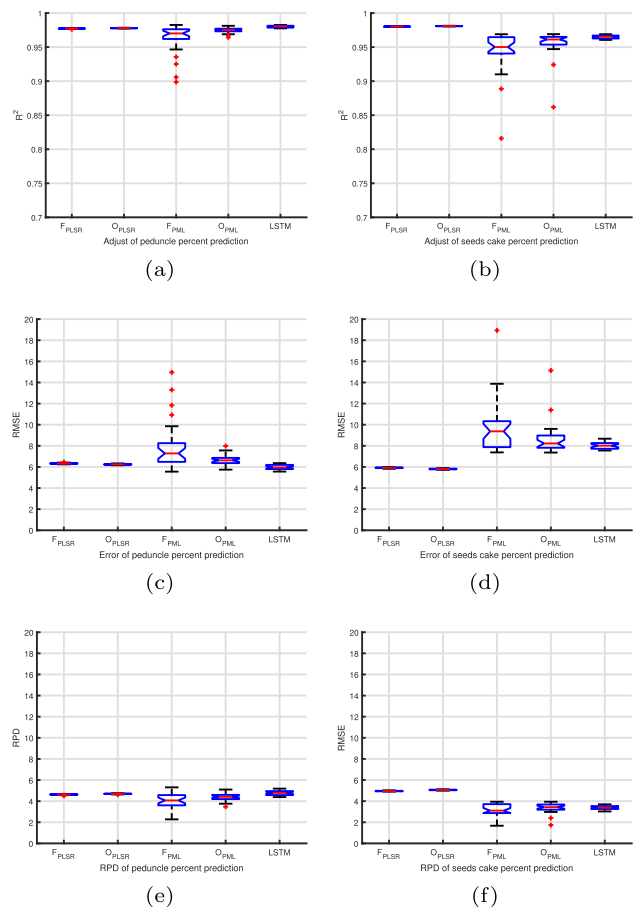


Fig. 9 Models’ metrics for adulterants prediction

3.5 Long short-term memory-based models

Finally, Fig. 8a and b shows the results of the regression model based on recurring neural networks, using the LSTM network. Results in Fig. 8a and b show the predictions of the powder of adulterated pericarp with stalk and seed cake.

The significance of comparing learning methods lies in the fact that each product exhibits different behaviors. In our case, when evaluating capsicum, where capsaicin and dihydrocapsaicin are the predominant compounds, we must not overlook the presence of other compounds that should be considered during the evaluation. Compounds such as colorants have been documented in these capsicum varieties [10], generating the need for prediction models to be tailored specifically to each model.

3.6 Models comparison

Predictive model results that full and optimized spectra profiles differences explain the feasibility of the NIRS to identify how adulterated pericarp powder with peduncle and seed cake are; then, Fig. 9 shows the box plot of each of the five analyzed models using the statistical metrics. According to Fig. 9a and b, the prediction models obtained R^2 higher than 0.96 or 0.90 for the content of the peduncle or seed cake, respectively. The prediction of the peduncle content showed a lower adjustment; mainly when relevant variables were used. Furthermore, fully optimized PLSR models present lower variability in R^2 compared to neural network-based models (Fig. 9a and b).

Evaluating RMSE of the PLSR optimized models, it obtained the lowest values for both prediction of the adulterant percentage, for peduncle powder (6.23) and seed cake (5.76); therefore, this model showed greater precision in the calibration set and classification. In the same way, Fig. 9c and d shows that the variability of the RMSE in neural network-based models PML and LSTM, respectively, was comparatively higher.

Finally, Fig. 9e and f RPD values showed all models could be considered reliable ($RPD > 2$); this was mainly true for PLSR models. However, this metric, which assesses the extent of error estimation compared to the standard deviation, exhibits significant variation for neural network-based models, which leads to the conclusion that models based on neural networks may not fit properly in these experimental settings. Other studies that evaluated models like those used here found that RMSE values were better when using neural networks; this occurs when working with samples with high humidity, such as corn [45], or with dry samples containing different active compounds, such as tea [43].

The optimized PLSR model showed the best predictive indicators in the study; this coincides with those reported

by [19] in their study of dry goods, [6] in onion powder. Both studies affirm that combining NIR with appropriate multivariate analysis can produce reliable results. These findings emphasize the need to create specific models for each product, considering the unique composition of the product under analysis. In the case of the Capsicum genus, the generation of capsaicin and dihydrocapsaicin oleoresins is specific to the pericarp, with the shape and composition differing in the peduncle and seeds, which could be the reason for differentiation in the NIR spectra.

4 Conclusions

In this paper, the feasibility of the estimation of the level of adulterant was studied in paprika powder. In the comparison, the proposed methodology was evaluated based on adulterated pericarp powder mixed with peduncle and seed cake powders using NIRS in conjunction with PLSR and neural network-based models. The most significant wavelengths for adulterant estimation were found within the range of 1600–2000 nm, with absorption bands at 1725 nm and 1761 nm. The aforementioned ranges are related to the functional group C=O, which is the most notable. In general, the predictive models based on PLSR outperformed the predictive methods based on neural networks, all of which have values R^2 higher than 0.95. However, based on RMSE and RPD values, optimized PLSR was shown to be the most effective among all predictive models. In conclusion, NIR spectrometry can be used to estimate the level of adulteration in paprika powder, when adulterants include peduncle or seed cake powder, and the highest performance was achieved by coupling it to PLSR when compared to models based on neuronal networks.

Further research may include the exploration of automatic methods for wavelength selection, to automate the whole process. It also may be interesting to explore other more sophisticated algorithms that involve higher dimensional representations of knowledge, at the expense of higher computational resources. Finally, it is still to be considered the application of the proposed methodology to different food products, which may include garlic, ginger, onion, black pepper, and other seasonings.

Acknowledgements The authors appreciate the support provided to this work by the project “Creation of the laboratory service for food safety research at the Universidad Nacional de Frontera,” with Unique Investment Code No. 2439545.

Data availability Data will be made available on reasonable request.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

- Choudhary A, Gupta N, Hameed F, Choton S (2020) An overview of food adulteration: concept, sources, impact, challenges and detection. *Int J Chem Stud* 8(1):2564–2573. <https://doi.org/10.22271/chemi.2020.v8.i1am.8655>
- Rodriguez-Saona L, Giusti M, Shotts M (2016) 4 - advances in infrared spectroscopy for food authenticity testing. In: Downey G (ed) *Advances in food authenticity testing*. Woodhead Publishing, Sawston, pp 71–116. <https://doi.org/10.1016/B978-0-08-100220-9.00004-7>
- de Lima A, Batista A, de Jesus J, Silva J, de Araújo A, Santos L (2020) Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling. *Food Control* 107:106802. <https://doi.org/10.1016/j.foodcont.2019.106802>
- Sasikumar B, Swetha VP, Parvathy VA, Sheeja TE (2016) 22 - advances in adulteration and authenticity testing of herbs and spices. In: Downey G (ed) *Advances in food authenticity testing*. Woodhead Publishing, Sawston, pp 585–624. <https://doi.org/10.1016/B978-0-08-100220-9.00022-9>
- Lee S, Lohumi S, Lim H, Gotoh T, Cho B, Kim M, Lee S (2015) Development of a detection method for adulterated onion powder using Raman spectroscopy. *J Fac Agric Kyushu Univ* 60(1):151–156. <https://doi.org/10.5109/1526312>
- Lohumi S, Lee S, Lee MW, Kim Mo C, Bae H, Cho B (2014) Detection of starch adulteration in onion powder by ft-nir and ft-ir spectroscopy. *J Agric Food Chem* 62(38):9246–9251. <https://doi.org/10.1021/jf500574m>
- Wilde A, Haughey S, Galvin-King P, Elliott C (2019) The feasibility of applying nir and ft-ir fingerprinting to detect adulteration in black pepper. *Food Control* 100:1–7. <https://doi.org/10.1016/j.foodcont.2018.12.039>
- Jäger M, Jiménez A, Amaya K (2013) *Guía de Oportunidades de Mercado Para Los Ajíes Nativos de Perú*. Bioersity International, Rome
- Collera-Zúñiga O, Jiménez F, Gordillo R (2005) Comparative study of carotenoid composition in three Mexican varieties of capsicum annum l. *Food Chem* 90(1):109–114. <https://doi.org/10.1016/j.foodchem.2004.03.032>
- Horn B, Esslinger S, Pfister M, Fahl-Hassek C, Riedl J (2018) Non-targeted detection of paprika adulteration using mid-infrared spectroscopy and one-class classification—is it data preprocessing that makes the performance? *Food Chem* 257:112–119. <https://doi.org/10.1016/j.foodchem.2018.03.007>
- Guillen N, Tito R, Mendoza N (2018) Capsaicinoids and pungency in capsicum Chinese and capsicum Baccatum fruits. *Pesqui Agropecu Trop* 48:237–244. <https://doi.org/10.1590/1983-40632018v48s2334>
- Van-Asselt ED, Banach JL, Van-Der-Fels-Klerx HJ (2018) Prioritization of chemical hazards in spices and herbs for European monitoring programs. *Food Control* 83:7–17. <https://doi.org/10.1016/j.foodcont.2016.12.023>
- Choudhary N, Sekhon B (2011) An overview of advances in the standardization of herbal drugs. *J Pharm Educ Res* 2(2):55
- Smillie T, Khan I (2010) A comprehensive approach to identifying and authenticating botanical products. *Clin Pharmacol Ther* 87(2):175–186. <https://doi.org/10.1038/clpt.2009.287>
- Aiello D, Siciliano C, Mazzotti F, Di Donna L, Athanassopoulos CM, Napoli A (2020) A rapid maldi ms/ms based method for assessing saffron (*Crocus sativus* L.) adulteration. *Food Chem* 307:125527. <https://doi.org/10.1016/j.foodchem.2019.125527>
- Zalacain A, Ordoudi S, Díaz-Plaza E, Carmona M, Blázquez I, Tsimidou GM, Alonso GL (2005) Near-infrared spectroscopy in saffron quality control: Determination of chemical composition and geographical origin. *J Agric Food Chem* 53(24):9337–9341. <https://doi.org/10.1021/jf050846s>
- Wilson A (2013) Diverse applications of electronic-nose technologies in agriculture and forestry. *Sensors* 13(2):2295–2348. <https://doi.org/10.3390/s130202295>
- Aykas DP, Menevseoglu A (2021) A rapid method to detect green pea and peanut adulteration in pistachio by using portable ft-mir and ft-nir spectroscopy combined with chemometrics. *Food Control* 121:107670. <https://doi.org/10.1016/j.foodcont.2020.107670>
- Wang Z, Wu Q, Kamruzzaman M (2022) Aportable nir spectroscopy and pls based variable selection for adulteration detection in quinoa flour. *Food Control* 138:108970. <https://doi.org/10.1016/j.foodcont.2022.108970>
- Boadu VG, Teye E, Amuah CL, Lamptey FP, Sam-Amoah LK (2023) Portable nir spectroscopic application for coffee integrity and detection of adulteration with coffee husk. *Processes* 11(4):1140. <https://doi.org/10.3390/pr11041140>
- Couto C, Souza Coelho C, Moraes Oliveira EM, Casal S, Freitas-Silva O (2023) Adulteration in roasted coffee: a comprehensive systematic review of analytical detection approaches. *Int J Food Prop* 26(1):31–258. <https://doi.org/10.1080/10942912.2022.2158865>
- Galvin-King P, Haughey SA, Elliott CT (2020) The detection of substitution adulteration of paprika with spent paprika by the application of molecular spectroscopy tools. *Foods* 9(7):944. <https://doi.org/10.3390/foods9070944>
- Zhao X, Wang Y, Liu X, Jiang H, Zhao Z, Niu X, Li C, Pang B, Li Y (2022) Single- and multiple-adulterants determinations of goat milk powder by NIR spectroscopy combined with chemometric algorithms. *Agriculture* 12(3):434. <https://doi.org/10.3390/agriculture12030434>
- Xu Y (1996) Tutorial: capillary electrophoresis. *Chem Educ* 1(2):1–14. <https://doi.org/10.1007/s00897960023a>
- Yip P, Chau C, Mak C, Kwan H (2007) DNA methods for identification of Chinese medicinal materials. *Chinese Medicine* 2(1):1–19. <https://doi.org/10.1186/1749-8546-2-9>
- Rinaldi C (2007) Authentication of the Panax Genus Plants Used in Traditional Chinese Medicine (TCM) Using Randomly Amplified Polymorphic DNA (RAPD) Analysis
- Xue C, Xue H, Li D (2009) Authentication of the traditional Chinese medicinal plant Saussurea involucre using enzyme-linked immunosorbent assay (elisa). *Planta Med* 75(09):15. <https://doi.org/10.1055/s-0029-1234669>
- Heidarbeigi K, Mohtasebi S, Foroughirad A, Ghasemi-Varnamkhaesti M, Rafiee S, Rezaei K (2015) Detection of adulteration in saffron samples using electronic nose. *Int J Food Prop* 18(7):1391–1401
- Blanco M, Villarroya I (2002) Nir spectroscopy: a rapid-response analytical tool. *TrAC Trends Anal Chem* 21(4):240–250. [https://doi.org/10.1016/S0165-9936\(02\)00404-1](https://doi.org/10.1016/S0165-9936(02)00404-1)
- Roggo Y, Chalou P, Maurer L, Lema-Martinez C, Edmond A, Jent N (2007) A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J Pharm Biomed Anal* 44(3):683–700. <https://doi.org/10.1016/j.jpba.2007.03.023>
- Li S, Zhang X, Shan Y, Su D, Ma Q, Wen R, Li J (2017) Qualitative and quantitative detection of honey adulterated with high-fructose corn syrup and maltose syrup by using near-infrared spectroscopy. *Food Chem* 218:231–236. <https://doi.org/10.1016/j.foodchem.2016.08.105>
- Castro W, Mejía J, De-la-Torre M, Acevedo-Juárez B, Bruno Tech AR, Avila-George H (2023) Radial grid reflectance correction for hyperspectral images of fruits with rounded surfaces. *Comput Electron Agric* 213:108179. <https://doi.org/10.1016/j.compag.2023.108179>
- Yoplac I, Avila-George H, Vargas L, Robert P, Castro W (2019) Determination of the superficial citral content on microparticles: an

- application of NIR spectroscopy coupled with chemometric tools. *Heliyon* 5(7):02122. <https://doi.org/10.1016/j.heliyon.2019.e02122>
34. Zhao Y, Wang S-H, Li Z, Cao F, Pei Z (2016) A novel interval integer genetic algorithm used for simultaneously selecting wavelengths and pre-processing methods. *Chin J Anal Chem* 44(9):1609–1616. [https://doi.org/10.1016/S1872-2040\(16\)60928-3](https://doi.org/10.1016/S1872-2040(16)60928-3)
 35. Shariati-Rad M, Hasani M (2010) Selection of individual variables versus intervals of variables in PLSR. *J Chemom J Chemom Soc* 24(2):45–56. <https://doi.org/10.1002/cem.1266>
 36. Pereira A, Reis M, Saraiva P, Marques J (2011) Development of a fast and reliable method for long-and short-term wine age prediction. *Talanta* 86:293–304. <https://doi.org/10.1016/j.talanta.2011.09.016>
 37. Castro W, De-la-Torre M, Avila-George H, Torres-Jimenez J, Guivin A, Acevedo-Juárez B (2022) Amazonian cacao-clone nibs discrimination using NIR spectroscopy coupled to naïve bayes classifier and a new waveband selection approach. *Spectrochim Acta Part A Mol Biomol Spectrosc* 270:120815. <https://doi.org/10.1016/j.saa.2021.120815>
 38. Prakash M, Sarin J, Rieppo L, Afara I, Töyräs J (2017) Optimal regression method for near-infrared spectroscopic evaluation of articular cartilage. *Appl Spectrosc* 71(10):2253–2262. <https://doi.org/10.1177/0003702817726766>
 39. Vásquez N, Magán C, Oblitas J, Chuquizuta T, Avila-George H, Castro W (2018) Comparison between artificial neural network and partial least squares regression models for hardness modeling during the ripening process of swiss-type cheese using spectral profiles. *J Food Eng* 219:8–15. <https://doi.org/10.1016/j.jfoodeng.2017.09.008>
 40. Wu L, Liu Z, Bera T, Ding H, Langley DA, Jenkins-Barnes A, Furlanello C, Maggio V, Tong W, Xu J (2019) A deep learning model to recognize food contaminating beetle species based on elytra fragments. *Comput Electron Agric* 166:105002
 41. Chen C, Yang B, Si R, Chen C, Chen F, Gao R, Li Y, Tang J, Lv X (2021) Fast detection of cumin and fennel using NIR spectroscopy combined with deep learning algorithms. *Optik* 242:167080. <https://doi.org/10.1016/j.ijleo.2021.167080>
 42. Pang L, Wang L, Yuan P, Yan L, Yang Q, Xiao J (2021) Feasibility study on identifying seed viability of *Sophora japonica* with optimized deep neural network and hyperspectral imaging. *Comput Electron Agric* 190:106426. <https://doi.org/10.1016/j.compag.2021.106426>
 43. Hong Z, Zhang C, Kong D, Qi Z, He Y (2021) Identification of storage years of black tea using near-infrared hyperspectral imaging with deep learning methods. *Infrared Phys Technol* 114:103666. <https://doi.org/10.1016/j.infrared.2021.103666>
 44. Soni A, Al-Sarayreh M, Reis MM, Brightwell G (2021) Hyperspectral imaging and deep learning for quantification of clostridium sporogenes spores in food products using 1d-convolutional neural networks and random forest model. *Food Res Int* 147:110577. <https://doi.org/10.1016/j.foodres.2021.110577>
 45. Zhang L, Zhang Q, Wu J, Liu Y, Yu L, Chen Y (2022) Moisture detection of single corn seed based on hyperspectral imaging and deep learning. *Infrared Phys Technol* 125:104279. <https://doi.org/10.1016/j.infrared.2022.104279>
 46. ElMasry G, Nakauchi S (2016) Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality—a comprehensive review. *Biosyst Eng* 142:53–82. <https://doi.org/10.1016/j.biosystemseng.2015.11.009>
 47. Castro W, Oblitas J, Rojas EE, Avila-George H (2020) partial least square regression for food analysis: basis and example. In: *Mathematical and statistical applications in food engineering*, pp 141–160
 48. Agudelo-Cuartas C, Granda-Restrepo D, Sobral PJ, Castro W (2021) Determination of mechanical properties of whey protein films during accelerated aging: application of ftir profiles and chemometric tools. *J Food Process Eng* 44(5):13477
 49. Dai Q, Cheng J-H, Sun D-W, Pu H, Zeng X-A, Xiong Z (2015) Potential of visible/near-infrared hyperspectral imaging for rapid detection of freshness in unfrozen and frozen prawns. *J Food Eng* 149:97–104. <https://doi.org/10.1016/j.jfoodeng.2014.10.001>
 50. Wehling RL (2010) 23 - infrared spectroscopy. In: Nielsen SS (ed) *Food analysis*. Aspen Publishers, Gaithersburg, pp 407–419. https://doi.org/10.1007/978-1-4419-1478-1_23

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Wilson Castro¹ · Jimy Oblitas² · Luis Nuñez¹ · Ives Yoplac³ · Himer Avila-George⁴ · Miguel De-la-Torre⁴ 

✉ Miguel De-la-Torre
miguel.dgomora@academicos.udg.mx

Wilson Castro
wcastro@unf.edu.pe

Jimy Oblitas
jimyoblitas@unc.edu.pe

Luis Nuñez
lnunez@unf.edu.pe

Ives Yoplac
ives.yoplac@untrm.edu.pe

Himer Avila-George
himer.avila@academicos.udg.mx

¹ Facultad de Ingeniería de Industrias Alimentarias y Biotecnología, Universidad Nacional de Frontera, Av. San Hilarión No. 101, Sullana 20103, Piura, Peru

² Facultad de Ingeniería, Universidad Privada del Norte, Vía-Evitamiento s/n, Cajamarca 06002, Cajamarca, Peru

³ Laboratorio de Nutrición Animal y Bromatología de Alimentos, Facultad de Ingeniería-Zootecnista, Agronegocios y Biotecnología, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Calle Higos Urco NO 350, Chachapoyas 01001, Chachapoyas, Peru

⁴ Departamento de Ciencias Computacionales e Ingenierías, Universidad de Guadalajara, Carretera Guadalajara-Ameca km 45.5, 46600 Ameca, Jalisco, Mexico