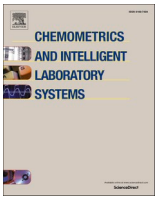
Chemometrics and Intelligent Laboratory Systems 250 (2024) 105137

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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



Automatic non-destructive estimation of polyphenol oxidase and peroxidase enzyme activity levels in three bell pepper varieties by Vis/NIR spectroscopy imaging data based on machine learning methods

Meysam Latifi Amoghin [a], Yousef Abbaspour-Gilandeh [a,**], Mohammad Tahmasebi [a], Juan Ignacio Arribas [b,c,d,*]

[a] Department of Biosystems Engineering, College of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran
[b] Castilla-Leon Neuroscience Institute, University of Salamanca, 37007, Salamanca, Spain
[c] Department of Electrical Engineering, University of Valladolid, 47011, Valladolid, Spain
[d] Artificial Intelligence Center, University of Valladolid, 47011, Valladolid, Spain

ARTICLE INFO

Keywords:
Effective wavelengths (EW)
Neural network
Non-destructive evaluation
Vis/NIR imaging spectroscopy
Polyphenol oxidase enzyme (PPO)
Peroxidase enzyme (POD)

ABSTRACT

The browning process of food products if often formed upon cutting and damage during their processing, transport, and storage, amongst other potential sources and reasons. Enzymic browning can be mainly due to polyphenol oxidase (PPO) and peroxidase (POD) enzymes. Visible/near-infrared (Vis/NIR) imaging spectroscopy in the range of 350–1150 nm was used in this study for automatic and non-destructive evaluation of PPO and POD activity levels in three bell pepper varieties (red, yellow, orange; N = 30), with a total of 30 inputs samples in each variety. The spectral data were then modeled by the partial least squares regression (PLSR) throughout the whole spectral range, without using any subset of the most effective wavelength (EW) values. Regression determination coefficient (R^2) values for the estimation (prediction) of POD enzyme activity levels were 0.794, 0.772, and 0.726 for red, yellow, and orange bell peppers, respectively, all over the validation set. At the same time, the activity levels of PPO enzyme over bell peppers showed R^2 values of 0.901, 0.810, and 0.859, for red, yellow, and orange bell peppers, respectively, all over the validation set. In addition, a combination of support vector machine (SVM) with either genetic algorithms (GA), particle swarm optimization (PSO), ant colony optimization (ACO), or imperialistic competitive algorithms (ICA) hybrid machine learning (ML) techniques were used to select the optimal (discriminant) spectral EW wavelength values, and regression performance was consistently improved, to judge from higher regression fit R^2 values. Either 14 or 15 EWs were computed and selected in order of their discriminative power using previously mentioned ML techniques. The hybrid SVM-PSO method resulted the best one in the process of selecting the most effective wavelength values (nm). On the other hand, three regression methods comprising PLSR, multiple least regression (MLR), and neural network (NN), were employed to model the SVM-PSO selected EWs. The ratio of performance to deviation (RPD), the R^2 and the root mean square error (RMSE), over the test set, for the non-linear NN regression method exhibited better results as compared to the other two regression methods, being closely followed by PLSR, and therefore NN regression method was selected as the best approach for modeling the most effective spectral wavelength values in this study.

1. Introduction

Bell pepper (*Capsicum annuum* L.) is an important vegetable commercial value, which is also commonly known as sweet pepper. It is rich in vitamins (A and C) and bioactive compounds [1,2] such as phenols, flavonoids, carotenoids, capsaicinoids, and for all those reasons is gaining commercial importance these days [3].

Quality and safety issues regarding food are among the major concerns of food suppliers and consumers. With the ever-increasing need for high-capacity production and processing, the food industry is facing

* Corresponding author. Department of Electrical Engineering, University of Valladolid, 47011, Valladolid, Spain.
** Corresponding author.
E-mail addresses: abbaspour@uma.ac.ir (Y. Abbaspour-Gilandeh), jarribas@tel.uva.es (J.I. Arribas).

https://doi.org/10.1016/j.chemolab.2024.105137
Received 9 November 2023; Received in revised form 1 May 2024; Accepted 2 May 2024
Available online 9 May 2024

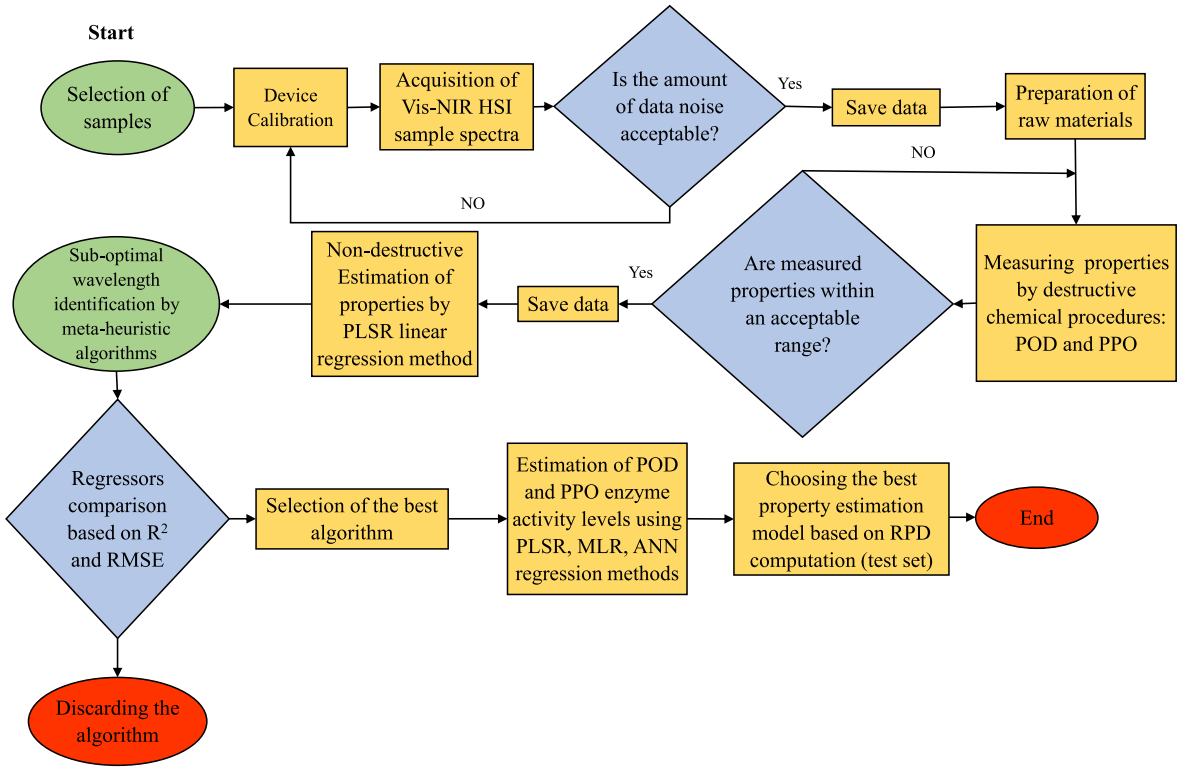


Fig. 1. Non-destructive estimation of polyphenol oxidase and peroxidase enzyme levels in pepper. System block diagram and implementation flowchart.

numerous new challenges. Quality management has the greatest impact on marketability. Food quality is generally defined by its composition and physical characteristics [4]. Although manual inspections are cost-effective for small-scale assessments, there is always the risk of overlooking critical quality and safety features. This implies the need for real-time, automatic, rapid, and non-destructive approaches to simultaneously quantify each of the quality parameters to ensure food safety.

Enzymatic browning is an oxidative-based natural process being an essential factor affecting food quality, such as its color, taste, and texture, while most fruits and vegetables are stored and processed [5]. Browning of food products may be caused by cuts, bruises, and damage during processing, transportation, and storage. Polyphenol oxidase (PPO) and peroxidase (POD) are responsible for enzymatic browning [6].

PPO is a copper-containing enzyme, catalyzing two reactions: hydrolysis produces cresolase activity, which converts monophenol into diphenol, and oxidation produces catecholase, which converts diphenol into a quinone molecule. Further polymerization of quinone may form a melanin compound [7].

Melanin is an either black, brown, or red pigment, responsible for the browning of the cut surface of the fruit/vegetable samples. POD is a heat-resistant oxidoreductase enzyme which causes browning in a wide variety of fruits and vegetables. It also catalyzes the oxidation of many phenolic compounds (substrates) which are naturally present in plants [8].

PPO and POD are responsible for enzymatic browning of agricultural products which has a negative impact on the commercial value of several crops such as cucumber [9], eggplant [10], and potato [11], among others.

Non-destructive automatic methods have been recently used to evaluate the quality of agricultural products, as these methods are suitable for online, reliable, and relatively low-cost inspection according to recent advances in sensor technology [12]. Some studies have addressed the use of near-infrared spectroscopy (NIR) to evaluate the qualitative characteristics of fruits and vegetables such as the prediction of the soluble solid content (SSC) in peach [13], the estimation of the maturity level of different types of fruits [14], the prediction of sugar content of citrus fruits [15], the non-destructive prediction of the total soluble solids content in strawberries [16], prediction of the properties, cultivar and geographical origin in lemon [17], or determining the internal qualitative characteristics of pomegranate [18].

Various studies have also addressed the non-destructive estimation of enzymatic activity of agricultural and food products. Baltacıoglu et al. [19] used Fourier transform infrared (FTIR) spectroscopy to examine the secondary structure and conformational change of mushroom polyphenol oxidase during heat treatment. For this purpose, FTIR spectroscopy and comparisons were used to assess the variations in enzyme activity during heat treatment at different combinations of power (60, 80, and 100 %), temperatures (20–60 °C), and time (0–30 min). Enzyme inactivation above 99 % was achieved in the range of 100 % at 60 °C for 10 min. FTIR studies showed distinct spectral changes after ultrasonication at 20 °C.

Nadafzadeh et al. [20] developed a computer vision system to evaluate the browning process of banana skin using genetic programming (GP) modeling to predict peroxidase and polyphenol oxidase enzymes activity. To check enzymatic browning in bananas, images of fruits were first taken at 25 °C for nine days. These images were then analyzed by digital image processing to predict and evaluate POD and PPO enzymes during the browning process of banana fruit skin. To this end, seventeen color parameters were extracted from each image as non-destructive parameters, then PPO and POD were experimentally measured. Finally, two equations were obtained by GP modeling, which can be used to predict and detect changes in the activity of PPO and POD enzymes during fruit storage.

Yang et al. [21] proposed a novel approach to use hyperspectral imaging (HSI) techniques with a weighted combination of spectral data and image features by fuzzy neural network (FNN) for real-time prediction of PPO activity in litchi pericarp. Litchi images were attained by a hyperspectral reflectance imaging system in the range of 400–1000 nm. To improve the prediction accuracy, a decision strategy was

developed based on the weighted combination of spectral data and image features, where the optimal weights were determined by FNN to better estimate PPO activity. The results showed a hybrid decision-making model as the best hybrid decision-making model for calibration and PPO activity estimation.

It is also known that the levels of activity of oxidation enzymes such as PPO and POD might be affected by the rate of respiration as well as the content of vitamin C, anthocyanin, and the type and number of phenolic compounds in the product [22,23]. Also, the amount of peroxidase activity in climacteric fruits is thought to be proportional to the volume of ethylene released by, and during the time of ethylene release, the peroxidase enzyme activity level increases [24].

To the best of our knowledge, no research has been carried out to date for the automatic non-destructive simultaneous prediction of the activity of both PPO and POD enzymes in bell pepper cultivars. The purpose of this research is to explore the applicability of a visible and near-infrared (vis-NIR) imaging spectroscopy method for the non-destructive estimation of the activity of these two PPO and POD enzymes in three bell pepper varieties.

A system block diagram including flowchart, is depicted in Fig. 1 summarizing the whole processes involved in this research study.

2. Materials and methods

2.1. Sample selection

In this study, three varieties of bell pepper including red (Pasarella RZ F1), yellow (Kaliroy RZ F1), and orange (Bachata RZ F1) bell peppers were purchased from local greenhouses. A total of 30 samples were selected from each cultivar with uniform size, shape, and color, free from any sign of mechanical or fungal issues. Before measurements, the samples were kept at 25 °C during 2 h to reach room temperature.

2.2. Non-destructive tests

2.2.1. Vis/NIR imaging spectroscopy

Vis/NIR imaging spectroscopic tests were performed using a PS-100 spectrometer (Apogee Instruments Inc., Logan, Utah, USA) equipped with a CCD detector, 2048 pixels, at a resolution of 1 nm with a halogen-tungsten light source in the wavelength range of 350–1150 nm. The spectrometer was connected to the computer through a USB port and the resulting spectra were recorded by the SpectraWiz software. Before spectrometry, dark and white spectra were first defined and saved as reference. In this way, the dark spectrum was first recorded by turning off light source; then, with the light source on, a standard Teflon disc was used to obtain the reference white spectrum. Spectroscopic measurements were performed at several different location points of each sample and the data were recorded.

2.3. Investigation of enzymatic properties

2.3.1. Preparation of enzyme extracts

For this purpose, 10 g of fruit pulp was transferred into a blender and completely homogenized. The resulting pulp was then added to a 20 ml of enzyme extract solution containing sodium phosphate buffer 0.4 M (pH = 6.5) containing 4 % (w/v) Polyvinylpyrrolidone (Merck, Germany) and 1 % (v/v) Triton X-100 (Merck, Germany) and thoroughly mixed using a vortex (Labtron LS-100, Iran). The resulting solution was centrifuged at 4 °C for 10 min and 4000 rpm (LISA 2.5L centrifuge AFI, France) and the resulting supernatant was used to assess the enzymic activity [25].

2.3.2. Polyphenol oxidase (PPO) activity

To determine the activity of PPO, 75 μl of enzyme extract was mixed with 3 ml of 0.05 M sodium phosphate buffer (pH = 6.5) containing (w/v) 0.05 M catechol (Merck, Germany). The control sample was prepared in a similar way by using water instead of enzyme extract. The absorbance value was then measured at 450 nm and 25 °C for 10 min using a spectrophotometer (NanoDrop™ OneC, Thermo Fisher Scientific, USA) in the kinetic mode. Enzyme activity was expressed as absorbance variations per minute per gram of sample [26].

2.3.3. Peroxidase (POD) activity

To assess POD activity, 500 μl of enzyme extract was mixed with 1 ml of 0.05 M sodium phosphate buffer (pH = 6.5). Then, to start the reaction, 1 ml of 0.05 M sodium phosphate buffer (pH = 6.5) containing 50 μl of hydrogen peroxide (H_2O_2) (Merck, Germany) at 1.5 % (v/v) and 1 ml of p-phenyldiamine (Merck, Germany) at 1 % (w/v) was added. The control sample was prepared in the same way by substituting the enzyme extract with water. The absorbance value was measured at 485 nm for 10 min at 25 °C using a spectrophotometer (NanoDrop™ OneC, Thermo Fisher Scientific, USA) under kinetic mode. Enzyme activity was expressed as absorbance changes per minute per gram of sample [27].

2.4. Data analysis

2.4.1. Partial least squares regression (PLSR) and seeking the most effective (optimal) wavelengths (EW)

The practical application of spectroscopic non-destructive methods with a full range of wavelengths is difficult as they are costly and time-consuming. Thus, it is often necessary to find the effective wavelengths and limit their range to the minimum possible level. In this research, first, a PLSR model was applied on full data. In order to validate the ability of the developed models, the input data were uniform randomly divided into two disjoint (empty intersection) subset categories: either calibration or validation, with a ratio of 70 %–30 % split of the total number of samples, respectively. For this purpose, the validation of regression (estimation) models was done by computing three well-known performance indices: the root mean square error (RMSE), the coefficient of determination (R^2), and the ratio of performance to deviation (RPD), as defined next in Eqs. (1)–(3):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(p_i - d_i)^2}{N}} \tag{1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(p_i - d_i)^2}{\sum_{i=1}^{N}(d_i - \overline{d})^2} \tag{2}$$

$$RPD = \frac{SD}{RMSE} = \sqrt{\frac{N}{N-1}\frac{\sum_{i=1}^{N}(d_i - \overline{d})^2}{\sum_{i=1}^{N}(p_i - d_i)^2}} \tag{3}$$

where $SD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(d_i - \overline{d})^2}$ is the standard deviation (SD) of measured data, RMSE the root mean square error between predicted and measured (reference) data as defined in (1), d_i are the reference (actual or measured) data values, p_i are the predicted data values, $\overline{d}$ is the average (mean) of the reference (measured) data values, and N is the total number of input sample components (either the number of reference or predicted components).

RPD was computed and the performance of the regression (estimation) system classified following the recommendation value ranges given by Chang et al. [28]. In principle, if RPD <1, the model is very weak and not accurate. If 1.4>RPD >1 the model is considered a weak one, if 1.8 > RPD >1.4 it is a suitable model and can be used for evaluation and prediction, if 2 >RPD>1.8 it is a considered a good model and is possible for quantitative accurate predictions, if 2.5 > RPD >2 it is very good model and finally if RPD>2.5 it is a great model.

2.4.2. Modeling based on the effective wavelengths (EWs)

PLSR, MLR, and non-linear NN regression models were implemented

Table 1
Activity levels of PPO and POD enzymes for three bell pepper varieties (absorbance/min g): red, yellow, and orange bell pepper.

Enzyme level	Pepper Variety	Mean	SD	Minimum	Maximum
POD	Red	71.597	4.273	58.789	78.181
	Yellow	67.460	14.150	8.300	80.400
	Orange	61.891	4.876	48.762	69.809
PPO	Red	1.4710	0.9380	0.3190	5.1310
	Yellow	1.0808	0.4279	0.5166	2.0749
	Orange	2.7620	1.6730	0.2840	8.0500

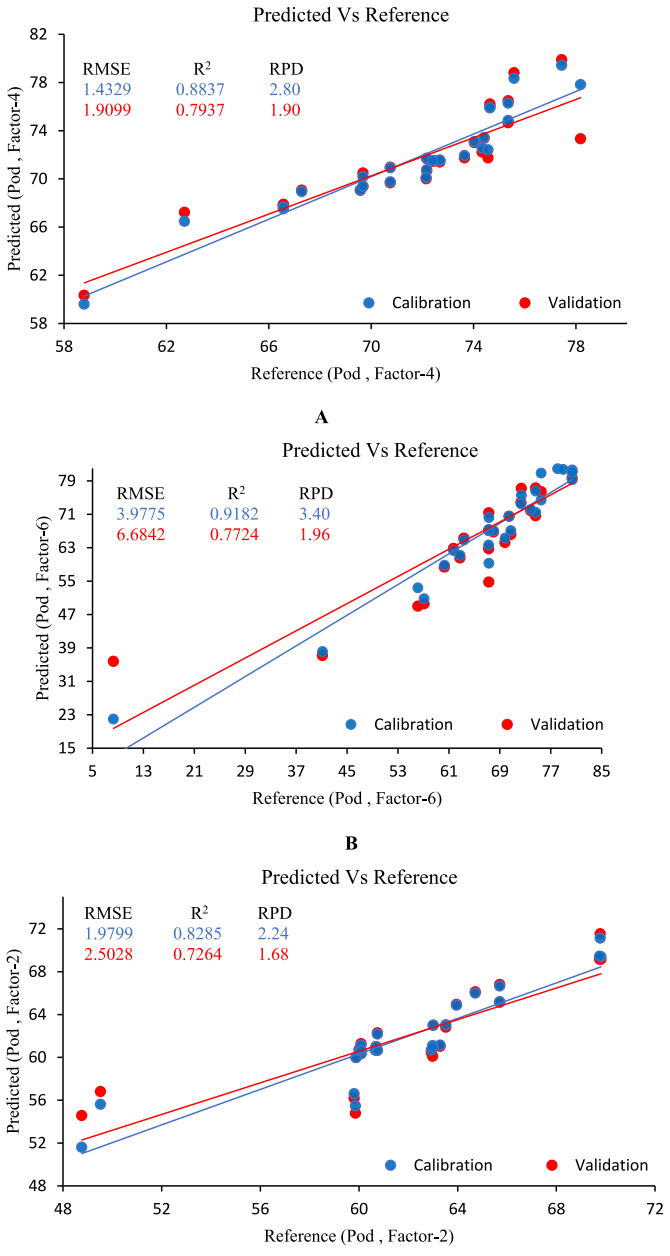


Fig. 2. Correlation between the actual and predicted values for calibration (blue dots) and validation (red dots) sets in estimation of POD enzyme activity based on the regression model of PLS; A) red, B) yellow, and C) orange pepper varieties: RMSE, R^2 and RPD performance indices computed. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

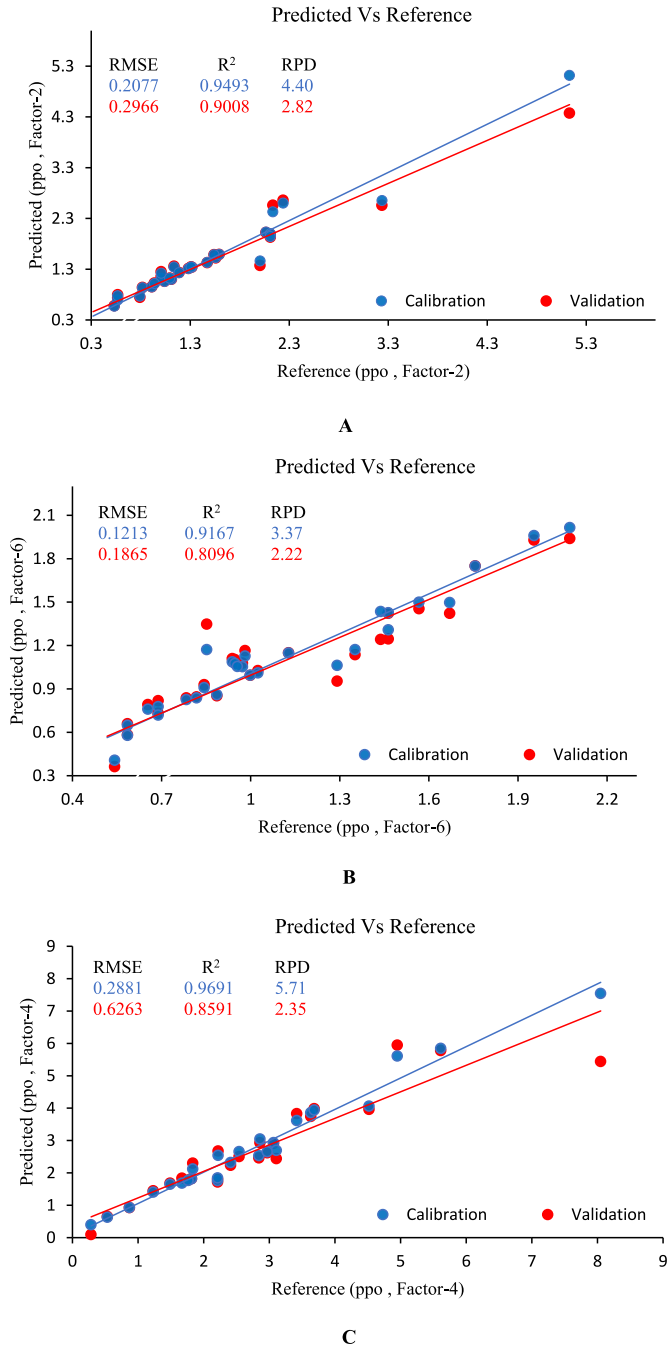


Fig. 3. Correlation between the actual and predicted values for calibration (blue dots) and validation (red dots) set in the estimation PPO enzyme activity based on the regression model of PLS: A) red, B) yellow, and C) orange pepper varieties: RMSE, R^2 and RPD performance indices computed. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

to find the best fitting model for the relationship between the EWs and the level of activity of PPO and POD enzymes in bell peppers. As mentioned in previous section, the statistical performance regression indices of RPD, R^2 and RMSE, see Eqs. (1)–(3), were employed to find the best fit model. A multilayer perceptron (MLP) neural network (NN) based on the well-known back-propagation learning algorithm including an input layer whose number of neurons is equal to the number of effective wavelengths and an output layer with one neuron (estimated activity of PPO and POD enzymes) and either one or two hidden layers with 10 neurons in each, was also used. The Levenberg-Marquardt

Table 2
EWs for POD based on various ML algorithms over the three bell pepper varieties under study.

Pepper Variety	Spectral Range (nm)	ML Methods	No. of EWs	Selected EWs (nm)
Red	580–980	SVM-GA	15	969 ،611 ،947 ،973.5 ،970 ،810.5 ،948.5 ،671.5 ،797 ،943 ،908 ،972 ،875.5 ،872 ،929
		SVM-PSO	15	917.5 ،952 ،668.5 ،980 ،973.5 ،967 ،972 ،942 ،950.5 ،956.5 ،975.5 ،929.5 ،967.5 ،979.5 ،963.5
		SVM-ACO	15	955.5 ،973.5 ،784 ،819 ،939 ،951 ،830.5 ،970 ،807 ،947.5 ،973 ،867 ،664.5 ،952 ،965.5
		SVM-ICA	14	972 ،959.5 ،931 ،849 ،954 ،977 ،967.5 ،729 ،782.5 ،979 ،713 ،974 ،940 ،920.5
Yellow	525–1000	SVM-GA	14	998.5 ،991 ،578 ،995 ،816 ،950 ،677 ،996.5 ،607.5 ،541 ،989 ،661.5 ،917.5 ،633.5
		SVM-PSO	15	988.5 ،1000 ،992.5 ،953.5 ،594 ،973 ،993 ،998.5 ،905.5 ،818.5 ،999 ،597.5 ،661 ،771 ،989.5
		SVM-ACO	15	998 ،536.5 ،667 ،998.5 ،993 ،974.5 ،908 ،999 ،811.5 ،906.5 ،601.5 ،820.5 ،990.5 ،965.5 ،981.5
		SVM-ICA	15	656.5 ،814 ،899.5 ،992 ،991 ،609 ،641.5 ،998.5 ،993 ،985 ،1000 ،997 ،550.5 ،733 ،991.5
Orange	565–960	SVM-GA	15	713 ،860 ،573.5 ،593.5 ،724.5 ،941 ،610.5 ،585.5 ،834.5 ،954.5 ،941.5 ،591.5 ،696.5 ،848.5 ،815.5
		SVM-PSO	15	612 ،940 ،578.5 ،754.5 ،750 ،943 ،742.5 ،592 ،803.5 ،939.5 ،585 ،858.5 ،571 ،593.5 ،942.5
		SVM-ACO	15	594 ،806 ،943.5 ،829.5 ،570 ،696 ،579 ،942.5 ،832.5 ،787.5 ،819 ،589.5 ،946.5 ،611.5 ،724
		SVM-ICA	15	579 ،759 ،575 ،773.5 ،941 ،582 ،782.5 ،890.5 ،598.5 ،940 ،608.5 ،952.5 ،602 ،796.5 ،812

Table 3
EWs of PPO based on various ML algorithms over the three bell pepper varieties under study.

Pepper Variety	Spectral Range (nm)	ML Methods	No. of EWs	Selected EWs (nm)
Red	580–980	SVM-GA	15	977 ،905 ،810 ،855 ،911.5 ،876.5 ،923.5 ،853.5 ،961.5 ،660.5 ،854 ،935.5 ،975.5 ،931 ،852.5
		SVM-PSO	15	745.5 ،952 ،972 ،976.5 ،956 ،956.5 ،975.5 ،953 ،815.5 ،807 ،871.5 ،950.5 ،907.5 ،942 ،913
		SVM-ACO	15	769 ،726.5 ،716 ،956 ،976.5 ،971.5 ،775 ،979 ،942 ،956.5 ،908 ،950.5 ،946 ،943 ،960.5
		SVM-ICA	15	605 ،816 ،931.5 ،942.5 ،974.5 ،887 ،976 ،899.5 ،848.5 ،952.5 ،926.5 ،901.5 ،850 ،807 ،952
Yellow	525–1000	SVM-GA	15	738.5 ،900 ،982 ،666.5 ،978.5 ،670 ،874 ،964 ،857.5 ،997.5 ،969 ،824.5 ،985 ،999 ،651.5
		SVM-PSO	14	761 ،981 ،984 ،960.5 ،952 ،663.5 ،982.5 ،997.5 ،689 ،941.5 ،991.5 ،995.5 ،649 ،976
		SVM-ACO	14	974.5 ،994.5 ،999 ،968.5 ،979 ،930 ،903.5 ،980 ،976.5 ،794 ،984.5 ،984 ،945.5 ،577
		SVM-ICA	15	855 ،973 ،682 ،853 ،646.5 ،970.5 ،989 ،996 ،729.5 ،999 ،983 ،971 ،982.5 ،988.5 ،638.5
Orange	565–960	SVM-GA	15	940.5 ،734.5 ،687.5 ،913 ،917.5 ،798 ،566.5 ،943.5 ،728 ،666.5 ،821 ،875.5 ،736.5 ،920 ،942
		SVM-PSO	14	936 ،941 ،921.5 ،570 ،942 ،913 ،914.5 ،949.5 ،852.5 ،918.5 ،931.5 ،920.5 ،941.5 ،919.5
		SVM-ACO	15	723.5 ،782 ،924.5 ،918 ،821.5 ،944 ،917.5 ،929.5 ،916.5 ،943.5 ،925.5 ،907 ،628.5 ،943 ،959
		SVM-ICA	15	605 ،941 ،621 ،877 ،668 ،917 ،944 ،919.5 ،908 ،922.5 ،623 ،593 ،642.5 ،943 ،922

algorithm was applied to update the weights of the MLP NN, which is one of the most common algorithms used in the field due to its fast training of the network and minimum error level. Data analysis was performed by Unscrambler X 10.4 (CAMO Software) and MatLab2022a software (MathWorks).

3. Results and discussion

The measured activity PPO and POD enzymes are shown next in Table 1 for the three bell pepper varieties.

3.1. Partial least squares regression (PLSR) for PPO and POD enzyme activity levels over the whole spectral range

The well-known PLSR linear regression method was employed to evaluate the correlation between the spectral data and the activity of either PPO or POD enzymes. The modeling was carried out on the full spectral data. Figs. 2 and 3, depict the correlation between the actual and predicted values of the POD and PPO enzymes activity using PLSR regression method. Accordingly, the R^2 values for the prediction of POD activity in red, yellow, and orange varieties were 0.794, 0.772, and 0.726, respectively and having 2>RPD>1.8, on average. For the prediction of PPO activity, the R^2 values were 0.901, 0.810, and 0.859, in red, yellow, and orange varieties, respectively and having 2.5>RPD>2, on average. The results show the ability of this methodology to accurately predict the activity levels of both peroxidase and polyphenol oxidase enzymes.

3.2. Optimization search of the most effective spectral wavelength (EW) values

Combinations of support vector machine (SVM) with either genetic algorithms (GA) [29], particle swarm optimization (PSO) [30], ant colony optimization (ACO) [31], or imperialistic competitive algorithm (ICA) [32] were adopted to find the sub-optimal EW values [33]. Either 14 or 15 wavelengths were extracted using each of the algorithms. Rajabi et al. [34] used a range of meta-heuristic algorithms to identify the optimal wavelengths for seed viability assessment, including world competitive contest (WCC), league championship algorithm (LCA), genetics (GA), PSO, ACO, ICA, learning automata (LA), heat transfer optimization (HTS), forest optimization (FOA), discrete symbiotic organisms search (DSOS), and cuckoo optimization (CUK). The results showed that all algorithms achieve a significant accuracy in predicting the allometric coefficient of seeds and reached correlation coefficients of more than 0.985 and errors below 0.0036, respectively.

Tables 2 and 3 show the effective wavelengths of POD and PPO enzymes based on different algorithms, respectively. Based on the main overtone bands [35], the bell pepper samples can be separated according to their enzymic activity in terms of the third OH overtone in wavelength around 770 nm and the second OH overtone in around 920 nm.

Two mean convergence performance indices were used to evaluate

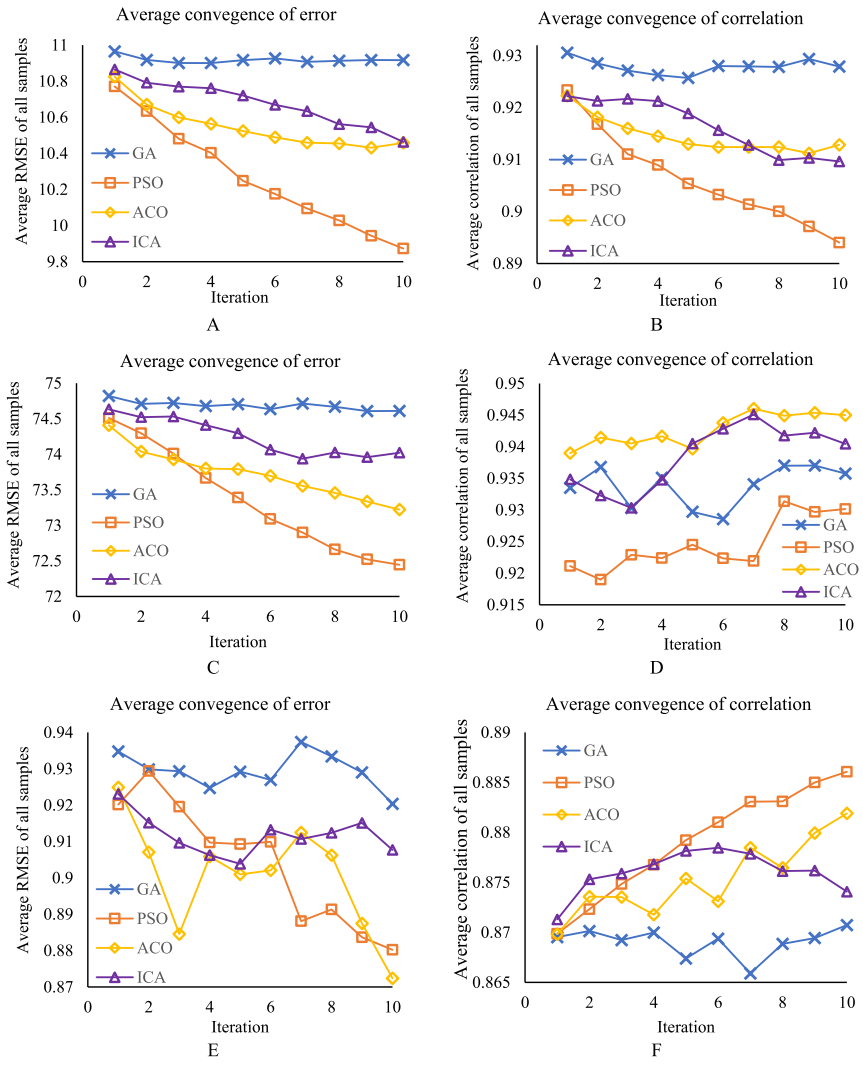


Fig. 4. Convergence diagrams produced from the dataset of POD activity over the three studied bell pepper varieties over all input samples using SVM. Diagrams compare the performance of algorithms based on both accuracy and error rate. RMSE and average correlation for A) and B) red, C) and D) yellow, and E) and F) orange, bell pepper varieties. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the optimized algorithms: the average RMSE and the average correlation of all samples. Average convergence implies that the results should improve by enhancing the number of iterations or the time allocated to the algorithms. Figs. 4 and 5, respectively, show the convergence diagrams produced using the combination of SVM method with GA, PSO, ACO, and ICA algorithms for predicting the POD and PPO activities over the three studied pepper varieties. These diagrams compare the performance of the algorithms in terms of accuracy and error score. Since the RMSE of the PSO and ACO algorithms showed a descending trend with increasing number of iterations and their average correlation is in a suitable range, it was concluded that these two algorithms show less error and more accuracy. On the other hand, as the time spent for analysis is one of the critical parameters, and the analysis must be achieved in the shortest time possible, given that the ACO algorithm takes more time, thus the PSO algorithm resulted more suitable in finding the effective wavelength values, online on real-time.

3.3. Optimal regression modeling using only the most effective wavelengths (EW) selected out of the whole spectral range

MLR, PLSR, and non-linear NN regression methods were employed to model the selected optimal EWs using the SVM-PSO algorithm.

Given that the use of non-destructive methods based on spectroscopy imaging with a full range of wavelengths might require longer acquisition time and additional computing resources, the practical application of this methodology introduces limitations and implementing data dimensionality reduction techniques to lessen computational complexity is often very desirable [36]. Therefore, one should seek to find a way to determine the EWs and limit the number of wavelengths used to the minimum possible value without a great reduction in accuracy. In order to reach the maximum sensitivity, spectroscopic measurements of absorption of substances should be performed at the wavelength corresponding to the absorption peak of that substance, since at this wavelength, absorbance per unit of concentration shows the largest changes and thus produces maximum accuracy. In this study, the PLSR method was used to model the spectral data in the full spectral range and in the range of EWs. The results showed that the RPD values of all cultivars (exception made of the red cultivar) in predicting POD enzyme activity in the range of EWs is higher than that in the full spectral range. Using the PLSR method in the range of EWs, simultaneously reduced the computed time and also increased the prediction (estimation) accuracy of the regression models.

Tables 4 and 5 list R^2, RMSE and RPD for the calibration and validation sets in PLSR and MLR regression models, respectively, based on effective wavelengths; at the same time, Table 6 shows R^2, RMSE and RPD values for training, validation, and test disjoint sets by the NN

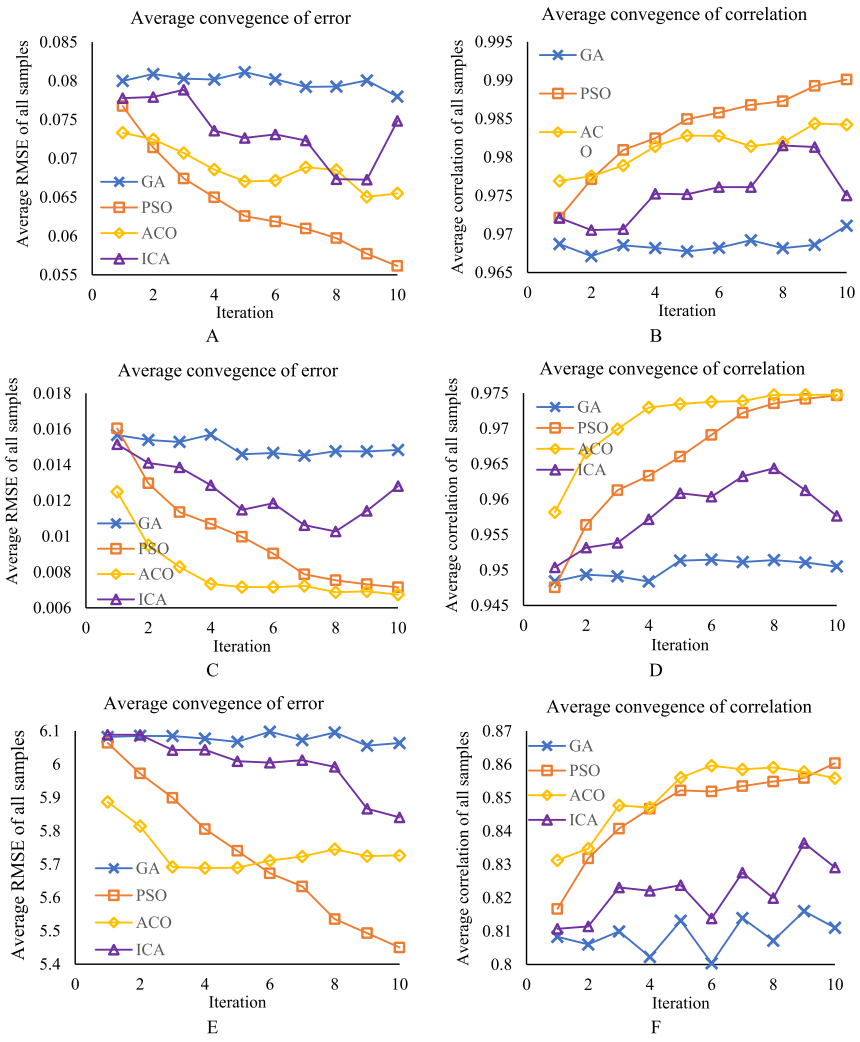


Fig. 5. Convergence diagrams produced from the dataset of PPO activity over the three studied bell pepper varieties for all input samples using SVM. Diagrams compare the performance of algorithms based on both accuracy and error rate. RMSE and average correlation for A) and B) red, C) and D) yellow, and E) and F) orange, bell pepper varieties. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 4
POD and PPO enzyme activity levels. R^2, RMSE and RPD regression values on three pepper varieties, over calibration and validation sets under PLSR regression model by using only effective (optimal) discriminative wavelengths (EWs). (cf. Tables 2 and 3).

Enzyme	Pepper Variety	Calibration			Validation		
		R^2	RMSE	RPD	R^2	RMSE	RPD
POD	Red	0.8334	1.7146	2.27	0.6606	2.5317	1.46
	Yellow	0.9105	4.1614	3.24	0.8097	6.2773	1.99
	Orange	0.8423	1.8980	2.35	0.7964	2.2433	1.95
PPO	Red	0.9427	0.2205	4.13	0.9230	0.2645	3.29
	Yellow	0.8769	0.1475	2.71	0.8463	0.1705	2.40
	Orange	0.9058	0.5032	3.16	0.8672	0.6214	2.53

Table 5
POD and PPO enzyme activity levels. R^2, RMSE and RPD regression values on three pepper varieties, over calibration and validation sets under MLR regression model by using only effective (optimal) discriminative wavelengths (EWs). (cf. Tables 2 and 3).

Enzyme	Pepper Variety	Calibration			Validation		
		R^2	RMSE	RPD	R^2	RMSE	RPD
POD	Red	0.9404	1.5013	2.76	0.6045	2.6870	1.86
	Yellow	0.9438	4.8251	2.84	0.6905	7.8704	1.78
	Orange	0.9422	2.1051	2.68	0.6132	2.4542	1.82
PPO	Red	0.9892	0.1398	6.67	0.6895	0.5225	1.41
	Yellow	0.9268	0.1608	2.56	0.6294	0.2604	1.60
	Orange	0.9705	0.4327	3.80	0.6690	0.9621	1.85

regression model. According to Tables 4 and 5, the R^2 value for the validation in PLSR method is higher than that in MLR method, and at the same time the RMSE value in PLSR method validation is lower than that in MLR. Also, by comparing the RPD values of PLSR and MLR models, it can be seen that the RPD value of the PLSR model is on average higher than that of the MLR model. Hence, the PLSR method offered better results here as compared to the MLR regression method. According to Table 6, the R^2, RMSE and RPD values for the NN non-linear method under the validation sub-set, exhibited better results as compared to the other two previously mentioned regression methods. Therefore, the NN method was selected as the best regression method for modeling effective wavelengths in POD and PPO enzyme activity levels estimation.

4. Conclusions

This work has examined the possibility of automatic, accurate, and

Table 6
POD and PPO enzyme activity levels. R^2, RMSE and RPD regression values on three pepper varieties, over training, validation, and test disjoint sets under Neural Network (NN) non-linear regression model using only effective (optimal) discriminative wavelengths (EWs). (cf. Tables 2 and 3).

Enzyme	Pepper Variety	Training			Validation			Test		
		R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD
POD	Red	0.9685	0.7013	3.49	0.9471	0.5865	3.71	0.9847	1.4443	2.81
	Yellow	0.9978	1.0577	4.23	0.9890	6.6937	3.96	0.9345	9.3358	2.70
	Orange	1	0	∞	0.7845	1.7038	2.95	0.9082	3.1085	3.13
PPO	Red	0.9898	0.0768	4.62	0.9643	0.0943	4.26	0.9996	0.5209	3.86
	Yellow	0.9930	0.0480	4.35	0.9799	0.0500	3.79	0.9012	0.0529	3.40
	Orange	1	0	∞	0.9886	0.2398	4.16	0.9886	0.2107	2.94

non-destructive evaluation (estimation) of both POD and PPO enzyme activity levels in three bell pepper varieties using Vis/NIR imaging spectroscopy. Absorbance spectroscopy was conducted in the wavelength range of 350–1150 nm. The full spectra range data were modeled without pre-processing using the PLSR method.

Given that the use of non-destructive methods based on imaging spectroscopy with a full range of wavelengths often need longer acquisition time and further resources, the practical application of this methodology introduces limitations, so a combination of support vector machine (SVM) with four machine learning algorithms (GA, PSO, ACO, and ICA) were proposed and employed to find the most effective (optimal) wavelengths (EWs). Each of these algorithms led to either 14 or 15 EWs, on average. Regarding the proper range of R^2 and RMSE values in the SVM-PSO method by increasing the number of iterations, this approach was selected as the winning algorithm to find the EWs.

Three regression methods were used and evaluated, MLR, PLSR, and non-linear NN, and were then used to model the EWs by SVM-PSO algorithm. To judge from results computed, the NN regression method shown the best results in modeling the consistent selection of EW values.

We believe that the difference in the activity of oxidation enzymes in the various pepper cultivars here considered might be due to the differences in those cultivars in terms of respiration rate, ethylene production, as well as vitamin C content, amongst others, despite further investigation is needed to confirm previous hypothesis.

To our best knowledge, the main potential limitations of proposed approach have to do with both the spectral limited resolution of the hyperspectral camera used in the analysis and the limited number N of input samples dataset used, since increased N values would lead to statistically more valid and accurate numerical results.

Main achievement follows next to conclude: after learning phase took place over the train set, automatic (machine learning) and accurate (as compared with reference values measured by common destructive chemical methodologies) estimation of both polyphenol oxidase and peroxidase enzyme activity levels (oxidative-based natural browning processes known to play a key factor degrading fruits and vegetables quality while being stored and processed) were computed in three bell pepper varieties from Vis/NIR imaging spectroscopy data, over the disjoint (empty intersection) test set.

CRediT authorship contribution statement

Meysam Latifi Amoghin: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Yousef Abbaspour-Gilandeh:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Mohammad Tahmasebi:** Writing – review & editing, Resources. **Juan Ignacio Arribas:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

J.I. Arribas wants to thank the Spanish Ministry for Science, Innovation and Universities (MICINN), Agencia Estatal de Investigacion (AEI), as well as to the Fondo Europeo de Desarrollo Regional funds (FEDER, EU), under grant number PID2021-122210OB-I00, by MCIN/AEI/10.13039/501100011033 and "ERDF A way of making Europe", European Union, for partially funding this work.

References

[1] N.Y. Shotorbani, R. Jamei, R. Heidari, Antioxidant activities of two sweet pepper Capsicum annuum L. varieties phenolic extracts and the effects of thermal treatment, Avicenna J. Phytomed. 3 (2013) 25.
[2] C.E. Cortés-Estrada, T. Gallardo-Velázquez, G. Osorio-Revilla, E. Castañeda-Pérez, O.G. Meza-Márquez, M. del Socorro López-Cortez, D.M. Hernández-Martínez, Prediction of total phenolics, ascorbic acid, antioxidant capacities, and total soluble solids of Capsicum annuum L. (bell pepper) juice by FT-MIR and multivariate analysis, Lebensm. Wiss. Technol. 126 (2020) 109285.
[3] M.R. Loizzo, A. Pugliese, M. Bonesi, F. Menichini, R. Tundis, Evaluation of chemical profile and antioxidant activity of twenty cultivars from Capsicum annuum, Capsicum baccatum, Capsicum chacoense and Capsicum chinense: a comparison between fresh and processed peppers, LWT–Food Sci. Technol. 64 (2015) 623–631.
[4] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, Hyperspectral imaging–an emerging process analytical tool for food quality and safety control, Trends Food Sci. Technol. 18 (2007) 590–598.
[5] S. Dong, Y. Ma, Y. Li, Q. Xiang, Effect of dielectric barrier discharge (DBD) plasma on the activity and structural changes of horseradish peroxidase, Qual. Assur. Saf. Crop Foods 13 (2021) 92–101.
[6] S. Pipliya, S. Kumar, P.P. Srivastav, Inactivation kinetics of polyphenol oxidase and peroxidase in pineapple juice by dielectric barrier discharge plasma technology, Innovat. Food Sci. Emerg. Technol. 80 (2022) 103081.
[7] F. Taranto, A. Pasqualone, G. Mangini, P. Tripodi, M.M. Miazzi, S. Pavan, C. Montemurro, Polyphenol oxidases in crops: biochemical, physiological and genetic aspects, Int. J. Mol. Sci. 18 (2017) 377.
[8] F.S. Burnette, Peroxidase and its relationship to food flavor and quality: a review, J. Food Sci. 42 (1977) 1–6.
[9] A.R. Miller, T.J. Kelley, C.V. Mujer, Anodic peroxidase isoenzymes and polyphenol oxidase activity from cucumber fruit: tissue and substrate specificity, Phytochemistry 29 (1990) 705–709.
[10] T. Docimo, G. Francese, M. De Palma, D. Mennella, L. Toppino, R. Lo Scalzo, G. Mennella, M. Tucci, Insights in the fruit flesh browning mechanisms in Solanum melongena genetic lines with opposite postcut behavior, J. Agric. Food Chem. 64 (2016) 4675–4685.
[11] A. Sanchez-Ferrer, F. Laveda, F. Garcia-Carmona, Partial purification of soluble potato polyphenol oxidase by partitioning in an aqueous two-phase system, J. Agric. Food Chem. 41 (1993) 1219–1224.
[12] R. Pandiselvam, V. Prithviraj, M. Manikantan, A. Kothakota, A.V. Rusu, M. Trif, A. Mousavi Khaneghah, Recent advancements in NIR spectroscopy for assessing the quality and safety of horticultural products: a comprehensive review, Front. Nutr. 9 (2022) 973457.
[13] S. Liu, W. Huang, L. Lin, S. Fan, Effects of orientations and regions on performance of online soluble solids content prediction models based on near-infrared spectroscopy for peaches, Foods 11 (2022) 1502.

[14] S.S.A. Shah, A. Zeb, W.S. Qureshi, M. Arslan, A.U. Malik, W. Alasmary, E. Alanazi, Towards fruit maturity estimation using NIR spectroscopy, Infrared Phys. Technol. 111 (2020) 103479.
[15] S.-Y. Kim, S.-J. Hong, E. Kim, C.-H. Lee, G. Kim, Application of ensemble neural-network method to integrated sugar content prediction model for citrus fruit using Vis/NIR spectroscopy, J. Food Eng. 338 (2023) 111254.
[16] A.C. Agulheiro-Santos, S. Ricardo-Rodrigues, M. Laranjo, C. Melgão, R. Velázquez, Non-destructive prediction of total soluble solids in strawberry using near infrared spectroscopy, J. Sci. Food Agric. 102 (2022) 4866–4872.
[17] L. Ruggiero, C. Amalfitano, C. Di Vaio, P. Adamo, Use of near-infrared spectroscopy combined with chemometrics for authentication and traceability of intact lemon fruits, Food Chem. 375 (2022) 131822.
[18] R. Khodabakhshian, B. Emadi, M. Khojastehpour, M.R. Golzarian, A comparative study of reflectance and transmittance modes of Vis/NIR spectroscopy used in determining internal quality attributes in pomegranate fruits, J. Food Meas. Char. 13 (2019) 3130–3139.
[19] H. Baltacıoğlu, A. Bayındırlı, F. Severcan, Secondary structure and conformational change of mushroom polyphenol oxidase during thermosonication treatment by using FTIR spectroscopy, Food Chem. 214 (2017) 507–514.
[20] M. Nadafzadeh, S.A. Mehdizadeh, M. Soltanikazemi, Development of computer vision system to predict peroxidase and polyphenol oxidase enzymes to evaluate the process of banana peel browning using genetic programming modeling, Sci. Hortic. 231 (2018) 201–209.
[21] Y.-C. Yang, D.-W. Sun, N.-N. Wang, A. Xie, Real-time evaluation of polyphenol oxidase (PPO) activity in lychee pericarp based on weighted combination of spectral data and image features as determined by fuzzy neural network, Talanta 139 (2015) 198–207.
[22] O.C. Othman, Polyphenoloxidase and perioxidase activity during open air ripening storage of pineapple (Ananas comosus L.), mango (Mangifera indica) and papaya (Carica papaya) fruits grown in Dar es Salaam, Tanzan. J. Sci. 38 (3) (2012) 84–94.
[23] O.S. Hutabarat, H. Halbwirth, Polyphenol oxidase and peroxidase activity in apple: dependency on cultivar and fruit processing, IOP Conf. Ser. Earth Environ. Sci. 355 (1) (2019) 1–8.
[24] A. Sahraei Khosh Gardesh, F. Badii, M. Hashemi, A.Y. Ardakani, N. Maftoonazad, A.M. Gorji, Effect of nanochitosan based coating on climacteric behavior and postharvest shelf-life extension of apple cv. Golab Kohanz, Lebensm. Wiss. Technol. 70 (2016) 33–40.
[25] A.S.K. Gardesh, F. Badii, M. Hashemi, A.Y. Ardakani, N. Maftoonazad, A.M. Gorji, Effect of nanochitosan based coating on climacteric behavior and postharvest shelf-life extension of apple cv. Golab Kohanz, Lebensm. Wiss. Technol. 70 (2016) 33–40.
[26] S.D.S. Fernandes, C.A.d.S. Ribeiro, M.F. Raposo, R. Morais, A.M.M.B.d. Morais, Polyphenol oxidase activity and colour changes of 'Starking' apple cubes coated with alginate and dehydrated with air, Food Nutr. Sci. 2 (2011) 451–457.
[27] N.S. Terefe, K. Matthies, L. Simons, C. Versteeg, Combined high pressure-mild temperature processing for optimal retention of physical and nutritional quality of strawberries (Fragaria× ananassa), Innovat. Food Sci. Emerg. Technol. 10 (2009) 297–307.
[28] C.-W. Chang, D.A. Laird, M.J. Mausbach, C.R. Hurburgh, Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties, Soil Sci. Soc. Am. J. 65 (2001) 480–490.
[29] J. McCall, Genetic algorithms for modelling and optimisation, J. Comput. Appl. Math. 184 (2005) 205–222.
[30] F. Marini, B. Walczak, Particle swarm optimization (PSO). A tutorial, Chemometr. Intell. Lab. Syst. 149 (2015) 153–165.
[31] M. Dorigo, E. Bonabeau, G. Theraulaz, Ant algorithms and stigmergy, Future Generat. Comput. Syst. 16 (2000) 851–871.
[32] E. Atashpaz-Gargari, C. Lucas, Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition, IEEE Congr. Evolut. Comput., IEEE (2007) 4661–4667.
[33] Y. Masoudi-Sobhanzadeh, H. Motieghader, A. Masoudi-Nejad, FeatureSelect: a software for feature selection based on machine learning approaches, BMC Bioinf. 20 (2019) 1–17.
[34] M. Rajabi-Sarkhani, Y. Abbaspour-Gilandeh, A. Moinfar, M. Tahmasebi, M. Martínez-Arroyo, M. Hernández-Hernández, J.L. Hernández-Hernández, Identifying optimal wavelengths from visible–near-infrared spectroscopy using metaheuristic algorithms to assess peanut seed viability, Agronomy 13 (2023) 2939.
[35] L.S. Magwaza, U.L. Opara, H. Nieuwoudt, P.J. Cronje, W. Saeys, B. Nicolaï, NIR spectroscopy applications for internal and external quality analysis of citrus fruit—a review, Food Bioprocess Technol. 5 (2012) 425–444.
[36] F. Tao, H. Yao, Z. Hruska, Y. Liu, K. Rajasekaran, D. Bhatnagar, Use of visible–near-infrared (Vis-NIR) spectroscopy to detect aflatoxin B1 on peanut kernels, Appl. Spectrosc. 73 (2019) 415–423.