

Detection of Apple Taste Information Using Model Based on Hyperspectral Imaging and Electronic Tongue Data

Jingjing Liu,^{1,2,3*} Simeng Liu,¹ Sze Shin,³ Fulong Liu,¹ Tie Shi,¹
Chuang Lv,¹ Qi Qiao,¹ Hairui Fang,¹ Wenjuan Jiang,¹ and Hong Men^{1**}

¹College of Automation Engineering, Northeast Electric Power University,
Jilin, Jilin 132012, China

²Department of Computer Science and Bioimaging Research Center, University of Georgia,
Athens, Georgia 30602, the United States

³Biosensor National Special Laboratory, Key Laboratory for Biomedical Engineering of Education Ministry,
Department of Biomedical Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China

(Received November 20, 2019; accepted March 17, 2020)

Keywords: hyperspectral, electronic tongue, competitive adaptive reweighted sampling algorithm, particle swarm optimization

Taste is one of the most important criteria for evaluating the quality of apples. In this study, a correlation model using hyperspectral images and quantitative taste information was built for the nondestructive detection of apple taste information. Firstly, the images of 90 sets of Aksu apples were collected by a hyperspectral image system and quantitative values of taste information (sourness and sweetness) were measured using an SA-402B electronic tongue. Secondly, to overcome the difficulties in obtaining the most representative wavelengths, a competitive adaptive reweighted sampling (CARS) algorithm was proposed to remove redundant information in the hyperspectral data. Then, 43 characteristic wavelengths corresponding to sourness and 22 characteristic wavelengths corresponding to sweetness were selected. Finally, particle swarm optimization (PSO) was used to dynamically optimize the kernel parameters and penalty factors of support vector regression (SVR). A PSO–SVR prediction model based on characteristic wavelengths was established. Upon comparing the performance of the prescreening and postscreening models, results showed that the CARS–PSO–SVR model achieved better prediction for apple taste information, for which the correlation coefficients (R^2) of sourness and sweetness were 0.81 and 0.887, and the root mean square errors of the prediction set (RMSEP) were 0.03 and 0.018, respectively.

1. Introduction

Apples are a fruit loved by people all over the world. Consumer choices are affected by the sensory qualities of products, especially their taste.^(1,2) Recent studies have shown that sweetness and sourness are the most important taste factors of apples in terms of consumer preferences. Some studies showed that the prediction of apples' sweetness and sourness is

*Corresponding author: e-mail: jingjing_liu@neepu.edu.cn

**Corresponding author: e-mail: menhong@neepu.edu.cn

<https://doi.org/10.18494/SAM.2020.2715>

inadequate because sweetness and sourness perceptions are influenced by the comprehensive bias of the human taste perception system towards multiple components.^(3,4) Among these components, the main substances that determine sourness are malic acid, succinic acid, oxalic acid, tartaric acid, acetic acid, and citric acid,^(5,6) whereas sweetness is mainly determined by fructose, glucose, and sucrose.

However, traditional physical and chemical methods can only evaluate taste characteristics by detecting the content of a single substance contributing to the sourness or sweetness of apples, which is unrepresentative. An electronic tongue is a multisensor taste detection system that is capable of simulating the human taste system.⁽⁷⁾ Its output response signals can reflect the overall information of a sample, and the obtained taste information is characterized objectively and digitally.⁽⁸⁾ Therefore, the electronic tongue can be used to obtain quantitative data for apple taste, providing a foundation to construct a correlation model for the perception of apple taste characteristics.

Hyperspectral imaging is an emerging nondestructive optical testing technology. In addition to acquiring spatial information, it can also obtain spectral information of a sample tested in the ultraviolet to near-infrared spectral range. The analysis and processing of spectral data by an appropriate manifold learning algorithm can objectively and reliably reflect the overall information of the sample tested, which is characterized in a fast, nondestructive, and accurate manner. It has been widely used for testing the quality of seeds,^(9,10) drugs,^(11,12) the interior and exterior of fruit and vegetables,^(13,14) and meat.^(15,16) The hyperspectral imaging used in apple quality testing mainly focuses on species classification,^(17,18) damage detection,^(19,20) and internal component detection.^(21,22)

Sourness and sweetness perception are influenced by a comprehensive balance reflecting many sensory characteristics of apples and the composition and properties of the hydrocarbon organic matter obtained as taste information by hyperspectral imaging.^(23,24) However, there have been few applications of joint sensory techniques for apple quality evaluation, mainly because sensory methodologies require more time and specialized resources than instrumental characterizations. Therefore, a correlation model between hyperspectral data and taste quantification values of apples was established in this work. An SA-402B electronic tongue system and a hyperspectral sorter were used to obtain the quantitative taste information and hyperspectral information of apples, respectively. A Savitzky–Golay (S–G) smoothing filter was adopted to eliminate noise in the spectral data. Competitive adaptive reweighted sampling (CARS) was used to remove redundant information from hyperspectral data caused by nonspecificity (existing in hyperspectral wavebands), interaction sensitivity, and the complexity of the components. Finally, a particle swarm optimization–support vector regression (PSO–SVR) model based on all the wavelengths range and a CARS–PSO–SVR model based on characteristic wavelengths were constructed. The two models were composed to provide technical support for the rapid and nondestructive detection of taste information. The processes are shown in Fig. 1.

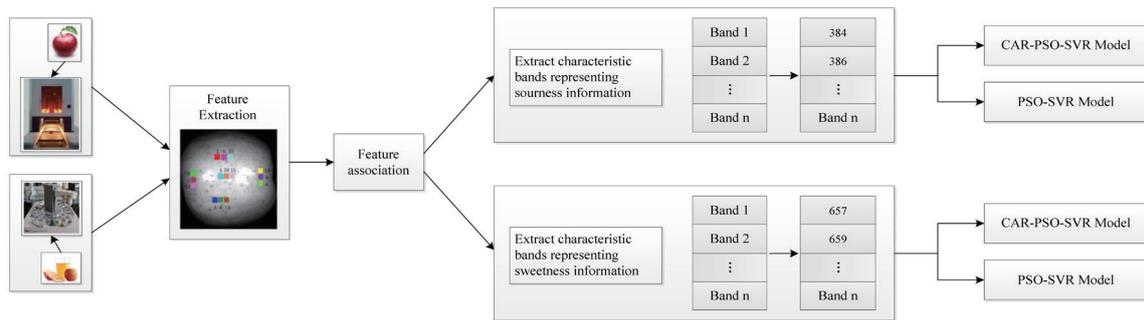


Fig. 1. (Color online) Processes of establishing a quantitative model fusing hyperspectral information and information from electronic tongue.

2. Materials and Methods

2.1 Sample selection and preparation

Aksu apples were selected as the research object. Apple samples with sizes ranging from 75 to 80 mm were purchased from a local market. All the apples were from the same batch. Each apple had a regular shape without bruising or other defects.

2.2 Hyperspectral image acquisition and correction

As shown in Fig. 2, a hyperspectral reflectance imaging system developed by Zolix was used to collect hyperspectral images of the apples. The core components of the instrument include four uniform light sources (25 W bromine tungsten lamp), a spectral camera (Image λ spectral image series, Zolix), an electronically controlled mobile platform, a computer (T440p, Lenovo Group), and control software (SpecView). The image size was 1344×1024 pixels and the spectral coverage ranged from 380 to 1038 nm in 256 bands.

The acquisition of hyperspectral images was controlled by SpecView software. Before the acquisition, the instrument parameters were adjusted to ensure that each acquired image was clear and undistorted. The exposure time of the camera was 15 ms, the target distance was 15 cm, and the speed of the mobile platform was 3 mm/s. The apples were placed directly below the spectral camera, and hyperspectral images of the apple samples were collected while moving the electronic control platform.

The hyperspectral image of each apple was corrected using a black moving platform and a white moving platform to eliminate the uneven distribution of light intensity in each band and the effect of the camera dark current. Hyperspectral images of a standard white calibration plate with 99% reflectance were collected, and a fully white calibration image R_w was obtained. The lens cap was then placed on the lens and a fully black calibration image R_d was obtained. The corrected hyperspectral image of each simple was calculated as

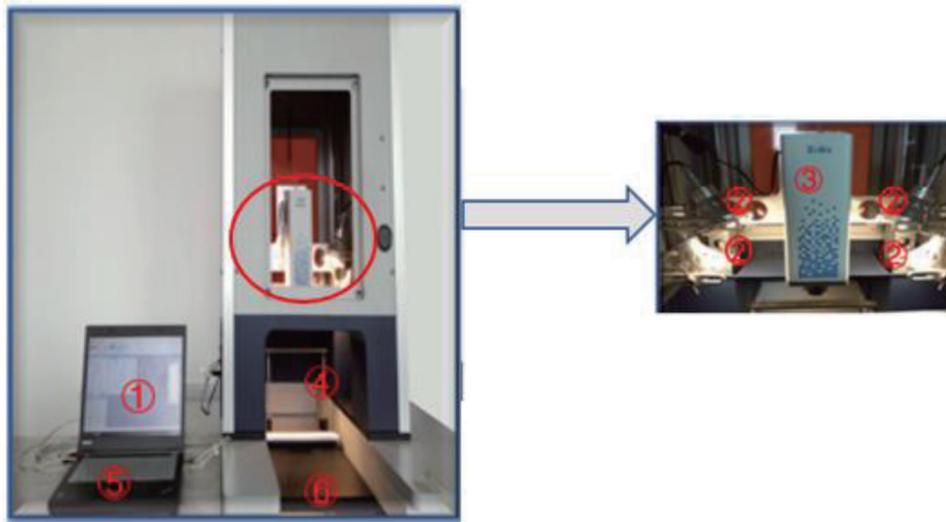


Fig. 2. (Color online) Hyperspectral reflectance imaging system: (1) SpecView control software, (2) four-light source (four 25 W bromine tungsten lamps), (3) spectral camera, (4) electronic camera, (5) computer, and (6) sample platform.

$$R = \frac{R_s - R_d}{R_w - R_d}, \quad (1)$$

where R_s is the original hyperspectral image, R is the corrected hyperspectral image, R_w is the fully white calibrated image, and R_d is the fully black calibrated image.

2.3 Taste information acquisition

The SA-402B electronic tongue system developed by Insent was adopted as the taste detection platform (Fig. 3). The system is based on the interaction between the taste cells on the tongue surface and the external stimuli in the actual tasting process. The detection results of the system can objectively and digitally represent the basic taste sensory indicators of the tested samples. The electronic tongue system includes a sensor array composed of working and reference electrodes, an automatic detection device, a data acquisition system, and a multifunctional numerical calculation platform.

In our experiment, a sensory array with two reference electrodes and two taste sensors was selected. The two reference electrodes were Ag electrodes containing an internal solution (3.33 mmol/L KCl + std AgCl).

After collecting the hyperspectral images of the apple samples, the apples were divided, juiced, and filtered, and 40 mL of supernatant was collected in a beaker. Then the supernatant needed to be diluted to meet the requirements of the e-tongue instrument. Prior to this, the sensors and reference electrodes were activated with a reference solution (30 mmol/L KCl + 0.3 mmol/L tartaric acid) and 3.3 mmol/L KCl solution, respectively. In this process, the apple samples, the positive and negative cleaning solutions, and reference solution were placed in



Fig. 3. (Color online) SA-402B e-tongue system: (1) measuring the aftertaste value, (2) fast cleaning for samples, (3) cleaning for positive and negative solutions, (4) positive and negative cleaning solutions, (5) for sensor calibration, (6) sensor reset solution, and (7) apple samples.

the reagent tank. The automatic detection device controlled the manipulator via the system parameters, ensuring that the positive and negative sensor array could be placed in contact with the sample solution, cleaning solution, and reference solution to clean the sensor and detect the sample. The hyperspectral average value for each apple was obtained by repeating the experiment three times. Finally, 90 sets of taste information were obtained by the electronic tongue system (sour taste data $Y_{90 \times 1}$, sweet taste data $Z_{90 \times 1}$).

2.4 Data processing method and model evaluation index

After collecting the spectral information of the apples, the redundant information in the spectral data was eliminated by S–G spectral pretreatment. Then, a quantitative prediction model based on the hyperspectral data and electronic tongue data of the apples was established by SVR. The correlation coefficient (R^2), root mean square error of the calibration set (RMSEC), and root mean square error of the prediction set (RMSEP) were selected as the indexes to optimize the parameters and evaluate the performance of the model. The range of R^2 is 0–1, and the closer R^2 is to 1, the higher the prediction accuracy of the quantitative prediction model. The smaller the values of RMSEC and RMSEP, the better the prediction effect of the model. The similarity between the two values indicates the stability of the model.

3. Data Preprocessing and Model Building

3.1 Spectral preprocessing and sample set division

Spectral noise is unavoidable in hyperspectral image acquisition and affects the data analysis. Smoothing filtering is one of the preprocessing methods commonly used in spectral

analysis. The effect of the S–G smoothing filter varies with the window width, enabling the filter to be used in many different situations.

In this experiment, S–G smoothing filtering was used to preprocess the original spectral data, thus eliminating data jitter caused by environmental temperature and humidity changes, and improving the prediction accuracy of the model. In our S–G smoothing filtering, when the width of the selection window was m , the width of the filter window was $n = 2m + 1$ and the measurement points were $x = (-m, -m + 1, \dots, 0, 1, \dots, m - 1, m)$, where the following polynomial of degree $k - 1$ was used to fit the data points in the window.

$$y = a_0 + a_1x + a_2x^2 + \dots + a_{k-1}x^{k-1} \quad (2)$$

If there are n such equations making up the k -element linear equations, for the equations to be valid, $n \geq k$ should be satisfied. Generally, $n > k$ was chosen, and fitting parameter A was determined by least squares fitting.

$$\text{Here, let } Y = \begin{bmatrix} y_{-m} \\ y_{-m+1} \\ \dots \\ y_m \end{bmatrix}, A = \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_{k-1} \end{bmatrix}.$$

Then,

$$\hat{A} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y. \quad (3)$$

The parameters of the S–G smoothing filtering method were set to $k = 2$ and $n = 15$. Five regions of interest (ROIs) were selected from each sample, and the average values were obtained. Then, the other sample regions were selected in turn. Finally, 90×256-dimensional spectral data of the 90 apple samples were obtained. Figure 4 shows the spectral curves of the 90 samples averaged over the ROIs after S–G pretreatment.

Using sample set partitioning based on point x – y distance (SPXY), the Kennard–Stone (K–S) algorithm was improved. The essence of the K–S algorithm is to select a set of spatially distributed datasets from the initial dataset as a training set, which was based on the Euclidean distance between sample points. The Euclidean distance equation is shown below:

$$d_x(\beta, \gamma) = \sqrt{\sum_{i=1}^k [x_\beta(i) - x_\gamma(i)]^2}, \beta, \gamma \in [1, P]. \quad (4)$$

Here $x_\beta(i)$ and $x_\gamma(i)$ are the reflectance of sample β and sample γ at the i th wavelength, respectively, K is the number of wavelengths, and $d_x(\beta, \gamma)$ is the distance between β and γ . The algorithm first selects the sample pairs (β, γ) corresponding to the largest $d_x(\beta, \gamma)$, then calculates the distances from the remaining sample to the reference points β and γ and selects the shorter distance from the reference points. Next, the sample corresponding to the maximum

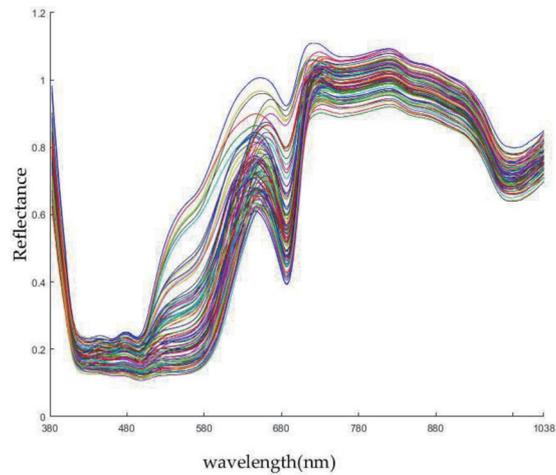


Fig. 4. (Color online) Spectral curves of apples after S–G pretreatment.

of these shortest distances is selected as a new reference point. Finally, this process is repeated until the specified number of samples. Similarly, when using SPXY, the detected object concentration was introduced as a feature in the distance calculation, and the spectral space and concentration space were merged. In this paper, the 90 sets of apple samples' hyperspectral data were divided into a training set and prediction set in the ratio of 2:1. Among them, the 60 sets of data in the training set were used to train the SVR model and the 30 sets of data in the prediction set were used to test the predictive ability and stability of the established SVR model. The specific training set and the parameters obtained from the prediction set are shown in Table 1.

3.2 Spectral data feature extraction

CARS is a variable selection method proposed to simulate Darwin's evolutionary theory of "survival of the fittest", which is helpful in finding feature bands that correspond to sourness and sweetness information in the hyperspectral data, which contain a large amount of redundant information. For each simulation, the absolute values of regression coefficients in the partial least squares (PLS) model are screened by the adaptive reweighted sampling (ARS) technique, and the wavelengths with small weights are removed. The minimum subset of the root mean square error of cross validation (RMSECV) in the PLS model is defined as the best variable.

In this paper, CARS was used to solve the problem of redundant data caused by the nonspecificity, interaction sensitivity, and complexity of components affecting apple taste information in hyperspectral bands. In this algorithm, Monte Carlo sampling (MCS) was first used to establish a PLS model between the corresponding sourness value and hyperspectral data $X_{(54 \times 256)}$. Secondly, an exponentially decreasing function (EDF) was used to remove the wavelengths that contributed least to the sourness prediction. Finally, the variables were further

Table 1

Quantitative statistics of apple taste information in training sets and prediction sets.

Sample type	Number of samples	Taste category	Range	Average value	Standard deviation
Training set	60	Sourness	-1.99–1.42	-0.73	0.63
		Sweetness	-16.82–28.3	-0.38	10.89
Prediction set	30	Sourness	-1.47–0.02	-0.82	0.38
		Sweetness	-15.8–16.59	-0.85	8.19

filtered by ARS. With the algorithm, an optimal combination of variables can be obtained by removing the wavelengths with small weights and selecting the minimum subset of the RMSECV through interactive validation. Then, before establishing the association model, the algorithm is used to select the feature set.

3.2.1 MCS

MCS is a random simulation method based on probability and statistical theory. Unlike other random classification algorithms, MCS can directly solve statistical problems involving continuous data without the need to discretize data in advance. Owing to the excellent performance of MCS, it was used to divide the hyperspectral data of apple samples into test sets and training sets without considering the linear relationship among them. The MCS also reduced the interference due to manual division, ensured accurate division, and saved time.

In the MCS algorithm, for statistically independent random variables x_i ($i = 1, 2, 3, \dots, k$), the corresponding probability density functions are $f(x_1), f(x_2), f(x_3), \dots, f(x_k)$. The function formula used in the algorithm is

$$z_i = g(x_1, x_2, \dots, x_k), i \in (1, 2, 3, \dots, N), \quad (5)$$

where $z_i \leq 0$. When $N \rightarrow \infty$, owing to Bernoulli's theorem of large numbers and the characteristics of normal random variables, the smaller the structural failure probability, the higher the reliability index.

Regardless of whether the state function is nonlinear or the random variable is non-normal, as long as the number of simulations is sufficient, a comparison can be obtained. Therefore, the number of samples was set to 50 in the process. Then 54 samples were extracted from the training set as a correction set (90%).

As the obtained data had multiple correlations between independent variables, the complexity of the multiple correlations should be reduced by finding the relationship between the independent variables and dependent variables. Thus, the PLS model can be used to maximize the covariance between the predictor (independent) matrix X and the predicted (dependent) matrix Y for each component of the reduced space.⁽²⁵⁾

In this algorithm, the PLS model is a method that models the covariance structure to find the basic relationship between two matrices, X and Y (hyperspectral data $X_{54 \times 256}$ and e-tongue sourness data $Y_{54 \times 1}$). In $X_{54 \times 256}$, 54 refers to the number of correction sets (90%) randomly

extracted from 60 training sets, and 256 is the number of hyperspectral bands acquired this time; $Y_{54 \times 1}$ represents the 54×1 -dimension sourness value matrix, the values of which are the sourness response values of the 54 calibration set samples we selected this time. The PLS model attempts to find the multidimensional direction of X to explain the multidimensional direction with the largest variance of Y , making it particularly suitable when the prediction matrix has more variables than the number observed and when there is multiple collinearity in the value of X .

In the PLS model, after using the R^2 matrix obtained by data normalization, the components of the independent variable group and the dependent variable group are selected separately. In this study, when the ratio of the current k components of the explanatory independent variables reaches 90%, the first k components are selected, then the following regression equation between the normalized index variable and the component variable of the k component pairs is found:

$$y = Xb + e. \quad (6)$$

Here, y is the sourness matrix with 54×1 dimensions, b is the coefficient matrix with 256×1 dimensions, and e is the predicted residual matrix with 54×1 dimensions. Furthermore, $b = [b_1, b_2, \dots, b_{256}]^T$, b_i ($1 \leq i \leq 256$) is the contribution of the i th wavelength to the sourness, and the larger the value of $|b_i|$, the greater the effect of the wavelength on the sourness.

Finally, the regression equation between the dependent variable group and the independent variable group is established, where the current k components are inserted in the regression equation obtained by Eq. (6). After obtaining the regression equation between the standardized index variables, the standardized regression variables are further converted into the original variables.

3.2.2 EDF

To select the EDF, for the j th sampling, the wavelength retention rate was calculated as

$$r_i = ae^{-kj}. \quad (7)$$

In Eq. (7), r_j denotes the retention rate of the j th sampling wavelength and a and k are constants satisfying two conditions. Firstly, all wavelengths are reserved in the first sampling operation; secondly, in the 50th sampling operation, two wavelengths are reserved. Then, Eqs. (8) and (9) are used to calculate a and k .

$$a = \left(\frac{p}{2}\right)^{1/(N-1)} \quad (8)$$

$$k = \frac{\ln(p/2)}{N-1} \quad (9)$$

Here, N is the number of samples and p is the number of bands. In this paper, N is 50 and p is 256. Therefore, we set $a = 1.104$ and $k = 0.1$.

3.2.3 ARS

We used ARS to filter the variables. Based on the principle of competitiveness, ARS introduces weight values and calculates the weight for each wavelength in the remaining wavelength set after removing the EDF. Variables were filtered by evaluating the weight of each wavelength. The weights w_i were calculated as

$$w_i = \frac{|b_i|}{\sum_{i=1}^{256} |b_i|}. \quad (10)$$

Finally, sampling was performed 50 times, and 50 sets of wavelengths were obtained from Eq. (10).

Figure 5(a) shows the number of characteristic wavelengths obtained from each sample. With increasing number of sampling operations in the presampling, the number of feature wavelengths decreased rapidly, but the speed of postsampling decreased. This indicates that the CARS model has two processes, rough selection and precise selection, in the selection of feature wavelengths. As can be observed from Fig. 5(b), when the number of samples increases,

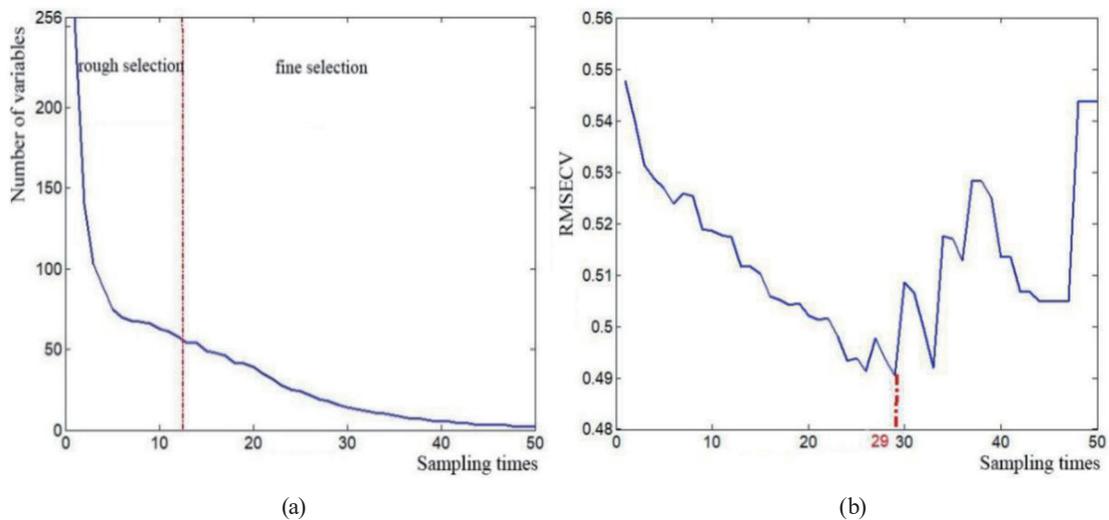


Fig. 5. (Color online) Results of using CARS to select the characteristic wavelengths corresponding to the sourness value: (a) feature variable selection; (b) trend of RMSECV.

the trend of the curve is a general decrease before 29 samples. The RMSECV value of the PLS model reached a minimum at 29 samples. This indicated that when 28 sampling operations were performed in the presampling, unrelated wavelengths were removed. After the 29th sampling, the RMSECV value increased gradually, indicating that the wavelengths associated with the apple sourness value were removed, resulting in poor forecast performance.

Therefore, the 29th sampling was selected as the predictive value of apple sourness, along with a total of 43 wavelengths. Similarly, we can obtain the characteristic wavelength corresponding to the sweetness value, because all feature wavelength values were obtained using the algorithm as shown in Table 2.

3.3 Establishment of correlation model and parameter analysis

After the characteristic wavelength was selected by CARS, the data sets corresponding to sourness ($X_{90 \times 43}$, $Y_{90 \times 1}$) and sweetness ($X_{90 \times 22}$, $Z_{90 \times 1}$) were separately obtained. MATLAB was used to construct the SVR model relating apple full-wavelength data, characteristic wavelength data, and apple taste information.

3.3.1 SVR model

SVR is an important branch of support vector machines (SVMs), and is an application model of SVMs in regression problems. The difference between SVR and SVM classification is that the SVM maximizes the distance of the sample point closest to the hyperplane, while SVR minimizes the distance of the sample point farthest from the hyperplane.

In this paper, an SVR model was selected to analyze sour and sweet tastes. The sample set was divided into 60 sets of apple spectral data with the corresponding taste data used as a training set to construct the model, and the remaining 30 samples were used to verify the predictive power and stability of the model. The data set corresponding to the sour taste ($X_{90 \times 43}$, $Y_{90 \times 1}$) was set to $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, $y \in [1, 90]$ and the data set corresponding to the sweet taste ($X_{90 \times 22}$, $Z_{90 \times 1}$) was set to $D = \{(x_1, z_1), (x_2, z_2), \dots, (x_m, z_m)\}$, $z \in [1, 90]$. These data were then input into the regression model for learning.

Taking hyperspectral data and sour taste information (x_m, y_m) as an example, SVR can accept a deviation between the predicted value and actual value, that is, if the absolute deviation

Table 2
Characteristic wavelength screening results of two types of taste information.

Taste information	Characteristic wavelength (nm)								
Sour	384	386	396	399	401	403	405	463	473
	475	485	490	500	509	524	527	530	537
	539	542	547	589	591	594	596	598	644
	647	652	662	669	672	684	687	689	730
	733	735	738	745	978	983	988		
Sweet	383	393	396	420	422	425	476	485	636
	657	659	662	702	705	707	940	943	946
	996	999	1002	1005					

between the predicted value and actual value is less than or equal to ε , the error is considered to be zero. However, when the absolute deviation is greater than ε , the calculation of the loss is equivalent to the construction of a 2ε -wide interval centered on $f(x)$. If the taste information value falls within the interval $[f(x) - \varepsilon, f(x) + \varepsilon]$, the model classification is considered correct, as shown in Fig. 6(a). Therefore, when establishing hyperspectral data and taste information models, the SVR model is robust and has good classification results.

In the SVR algorithm,

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \zeta(f(x_i) - y_i), \tag{11}$$

where C represents the regularization constant ζ . The calculation method of ζ is shown in Eq. (12).

$$\zeta(z) = \begin{cases} 0, & \text{if } |z| \leq \varepsilon \\ |z| - \varepsilon, & \text{otherwise} \end{cases} \tag{12}$$

By introducing relaxation variables ξ_i and $\hat{\xi}_i$, Eq. (11) can be rewritten as

$$\min_{w,b,\zeta,\hat{\xi}_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i), \tag{13}$$

such that

$$f(x_i) - y_i \leq \varepsilon + \xi_i,$$

$$y_i - f(x_i) \leq \varepsilon + \hat{\xi}_i,$$

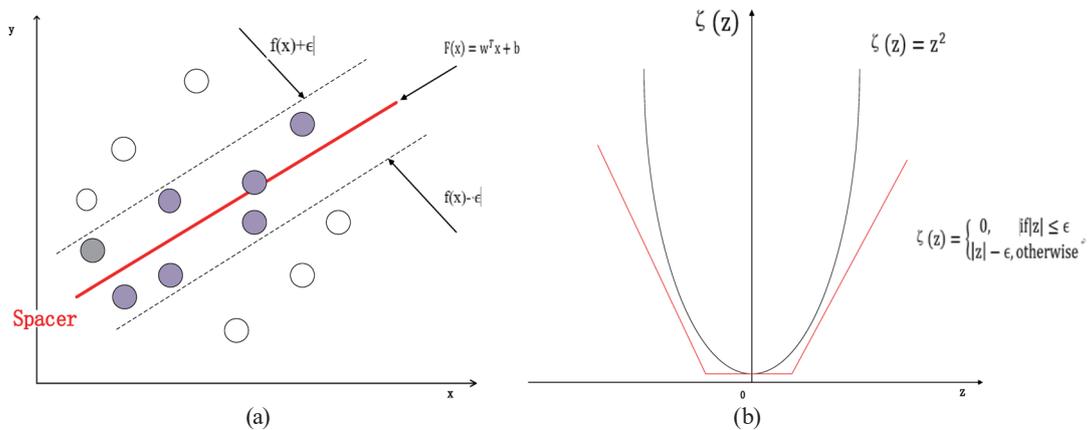


Fig. 6. (Color online) (a) SVR-related parameter. The red line shows the ε -spacer, and the samples lying between the dashed lines are not used to calculate the loss; (b) ε -insensitive loss function.

$$\xi_i \geq 0, \hat{\xi}_i \geq 0, i = 1, 2, \dots, m.$$

Similar to the SVM classification model, we can use the Lagrange function to transform the objective optimization function into an unconstrained form. The Lagrange function is constructed as

$$L(w, b, \xi, \hat{\xi}, \alpha, \hat{\alpha}, \mu, \hat{\mu}) = \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^m (\hat{\xi}_i + \xi_i) - \sum_{i=1}^m \hat{\mu}_i \hat{\xi}_i - \sum_{i=1}^m \hat{\mu}_i \xi_i + \sum_{i=1}^m (\hat{\mu}_i + \hat{\xi}_i) + \sum_{i=1}^m \alpha_i (f(x^{(i)}) - y^{(i)} - \varepsilon - \xi_i) + \sum_{i=1}^m \hat{\alpha}_i (y^{(i)} - f(x^{(i)}) - \varepsilon - \hat{\xi}_i). \quad (14)$$

Then the original function is transformed into the following dual function:

$$\min_{w, b, \xi, \hat{\xi}} \max_{\alpha, \hat{\alpha}, \mu, \hat{\mu}} L(w, b, \xi, \hat{\xi}, \alpha, \hat{\alpha}, \mu, \hat{\mu}) \Rightarrow \max_{\alpha, \hat{\alpha}, \mu, \hat{\mu}} \min_{w, b, \xi, \hat{\xi}} L(w, b, \xi, \hat{\xi}, \alpha, \hat{\alpha}, \mu, \hat{\mu}). \quad (15)$$

The minimum value of the optimization function can be obtained by the partial differentiation of the optimization function with respect to w , b , and ξ :

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^m x^{(i)} (\alpha_i - \hat{\alpha}_i), \quad (16)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^m (\alpha_i - \hat{\alpha}_i) = 0, \quad (17)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C - \alpha_i - \mu_i = 0, \quad (18)$$

$$\frac{\partial L}{\partial \hat{\xi}_i} = 0 \Rightarrow C - \hat{\alpha}_i - \hat{\mu}_i = 0. \quad (19)$$

Substituting the above four equations into Eq. (20) gives the dual form of the SVR model with only the parameter α :

$$\max \sum_{i=1}^m y^{(i)} (\hat{\alpha}_i - \alpha_i) - \varepsilon (\hat{\alpha}_i + \alpha_i) - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\hat{\alpha}_i - \alpha_i) (\hat{\alpha}_j - \alpha_j) k_{ij}, \quad (20)$$

such that

$$\sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) = 0,$$

$$0 < \alpha_i, \hat{\alpha}_i < C, i = 1, 2, \dots, m.$$

Karush–Kuhn–Tucker (KKT) conditions:

$$\begin{cases} \alpha_i(f(x^{(i)}) - y^{(i)} - \varepsilon - \xi_i) = 0, \\ \hat{\alpha}_i(y^{(i)} - f(x^{(i)}) - \varepsilon - \hat{\xi}_i) = 0, \\ (C - \alpha_i)\xi_i = 0, (C - \hat{\alpha}_i)\hat{\xi}_i = 0, \\ \alpha_i\hat{\alpha}_i = 0, \xi_i\hat{\xi}_i = 0. \end{cases} \quad (21)$$

Here, α_i can be nonzero if and only if $f(x^{(i)}) - y^{(i)} - \varepsilon - \xi_i = 0$ and $\hat{\alpha}_i$ can be nonzero if and only if $y^{(i)} - f(x^{(i)}) - \varepsilon - \hat{\xi}_i = 0$. In other words, only when the sample does not fall into the ε interval can the corresponding α_i and $\hat{\alpha}_i$ obtain nonzero values. In addition, the constraints $f(x^{(i)}) - y^{(i)} - \varepsilon - \xi_i = 0$ and $y^{(i)} - f(x^{(i)}) - \varepsilon - \hat{\xi}_i = 0$ cannot be satisfied simultaneously, so at least one of α_i and $\hat{\alpha}_i$ is zero. The solution of SVR is

$$f(x) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i)(x^{(i)})^T x + b. \quad (22)$$

The samples with $(\hat{\alpha}_i - \alpha_i) \neq 0$ are the support vectors of SVR, which must lie outside the interval. The basic idea of calculating α_i is to fix other parameters first, and then α_i is derived from other variables. These two steps are repeated until convergence. From Eq. (23), the value of b can be obtained as

$$b = y^{(i)} + \varepsilon - \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i)(x^{(i)})^T x. \quad (23)$$

All samples satisfying the conditions are used to solve b , and then the average value is calculated. The final solution of SVR is

$$f(x) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i)k(x, x^{(i)}) + b, \quad (24)$$

where $k(x, x^{(i)})$ is the kernel function, $\hat{\alpha}_i$ represents the Lagrangian multiplier, and b represents the offset.

The SVR model in Eq. (24) gives the relationship between the full-wavelength data of the apples, the characteristic wavelength data, and the apple taste information. In the model, the kernel function of SVR is RBF. The main factors affecting prediction accuracy are the parameters g of the RBF kernel function and the penalty factor c in the SVR model, PSO was used to select the optimal parameter combination (c, g) in this study.

3.3.2 Establishment of PSO and selection of model parameters

The PSO algorithm was used to correct the model of Hepper's simulated bird population (fish school) by allowing particles to fly to the solution space and land at the best solution, giving the algorithm its name. The advantage of PSO is that it is simple and easy to implement, does not require gradient information, and has few parameters. In particular, its coding characteristics for real numbers are particularly suitable for dealing with practical optimization problems.

In this optimization, the initial search ranges of parameters c and g in the PSO algorithm were set as $c \in [1.5, 100]$ and $g \in [1.7, 1000]$, and the number of iterations was set to 200. If (c, g) is regarded as a particle in the population, then the position and velocity of the particle are in two-dimensional space. Thus, supposing there are m particles in D -dimensional space, then the position and velocity of the m th particle can be expressed as $s_m = (s_{m1}, s_{m2}, s_{m3}, \dots, s_{mD})$ and $v_m = (v_{m1}, v_{m2}, v_{m3}, \dots, v_{mD})$, respectively.

A set of initial position points and velocity values was randomly assigned to each particle within a reasonable range. Then, the best location of particles was searched for with the purpose of minimizing the fitness function. The optimal solution and the global optimal were recorded as $p_g = (p_{g1}, p_{g2}, p_{g3}, \dots, p_{gD})$ and $p_m = (p_{m1}, p_{m2}, p_{m3}, \dots, p_{mD})$, respectively.

Accordingly, the position and velocity of particles were updated. Finally, the basic PSO formula was defined as

$$v_{mD}^{(n+1)} = wv_{mD}^{(n)} + c_1\delta_1(p_{mD}^{(n)} - s_{mD}^{(n)}) + c_2\delta_2(p_{mD}^{(n)} - s_{mD}^{(n)}), \quad (25)$$

$$s_{mD}^{(n+1)} = s_{mD}^{(n)} + v_{mD}^{(n+1)}, \quad m = 1, 2, \quad (26)$$

where n represents the current iteration number and w is the inertia weight, which adjusts the search ability of the solution space and whose value decreases linearly with time. c_1 and c_2 are learning factors, which are used to adjust the maximum step length of learning; in this study, we set c_1 to 1.5 and c_2 to 1.7. δ_1 and δ_2 are two random numbers in the range from 0 to 1, which are mainly used to increase the randomness of the search.

4. Results and Discussion

The PSO–SVR model in this study was constructed using 60 sets of apple spectral data taken from the sample sets and the corresponding taste data as training sets. The remaining 30 sample sets were used as test sets to verify the predictive ability and stability of the model. As shown in Table 3, the R^2 values of sourness and sweetness in the PSO–SVR model based on all the wavelengths (256 wavelengths) were 0.99 and 0.754, and the root mean square errors were 0.0017 and 0.047, respectively. In comparison, in the PSO–SVR model based on the characteristic wavelength, the R^2 of sourness and sweetness were 0.91 and 0.827, and the root mean square errors were 0.017 and 0.042, respectively.

Table 3
Comparison of prediction results between PSO–SVR and CARS–PSO–SVR models.

Model		Parameters		Training set		Testing set		Calculation time (s)
		c	g	R^2	RMSEC	R^2	RMSEP	
Full-wavelength	Sourness	1.669	641.4	0.99	0.0017	0.655	0.072	79
	Sweetness	100	69.75	0.754	0.047	0.59	0.104	116
Characteristic-wavelength	Sourness	3.903	1000	0.91	0.017	0.81	0.03	39
	Sweetness	100	1000	0.827	0.042	0.887	0.018	78

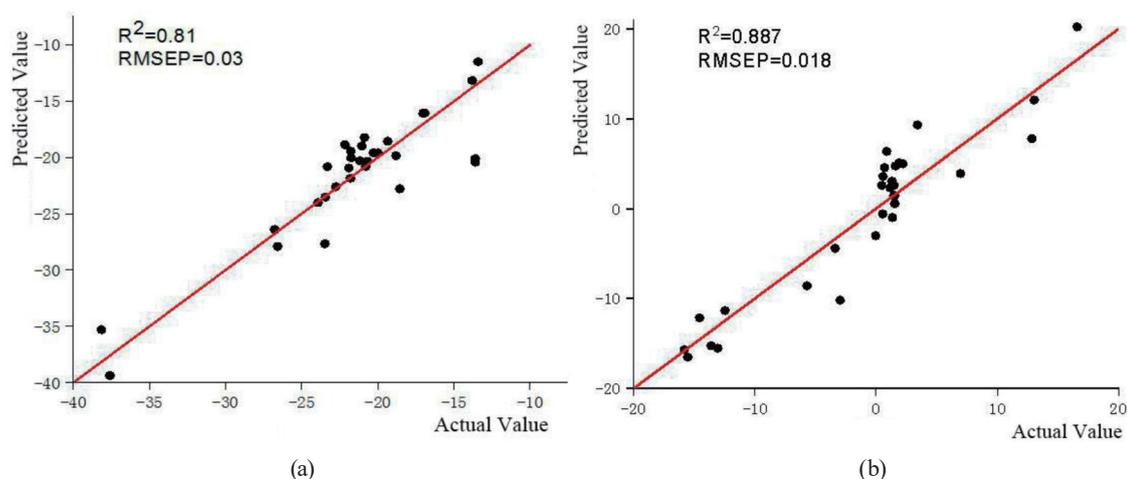


Fig. 7. (Color online) Prediction results of sourness and sweetness using regression model based on CARS-PSO-SVR. (a) Comparison of predicted value of sourness with the true value. (b) Comparison of the predicted value of sweetness with the true value.

The sourness and sweetness of apples were predicted by the PSO-SVR model, and the prediction results of the prescreening and postscreening for the characteristic variables obtained by the CARS model were compared and analyzed. It can be seen from Table 3 that after determining the optimal parameters c and g , for the prediction set of the CARS-PSO-SVR model, the R^2 of the sourness was 0.81, R^2 of the sweetness was 0.887, the root mean square error was 0.03, and the calculation time was 39 s, which were better than the values for the PSO-SVR model of sourness. For the training sets and prediction sets of the CARS-PSO-SVR model, the R^2 values of sweetness were 0.827 and 0.887, respectively, the root mean square error of the training set was 0.042, that of the prediction set was 0.018, and the calculation time was 78 s, which were all better than those for the PSO-SVR model of sweetness. The results showed that the CARS model eliminated the redundant information in the original spectral data, improved the prediction accuracy of the model, and shortened the calculation time. The prediction results are shown in Fig. 7. The R^2 values of the CARS-PSO-SVR model for sourness and sweetness were close to 1, and the differences between the RMSEC and RMSEP were small, indicating that the regression model had high accuracy and stability in predicting apple taste information.

5. Conclusions

In this paper, Aksu apples were used to explore a fast and nondestructive method of detecting apple taste information based on hyperspectral technology. Ninety sets of hyperspectral images were obtained by using visible-near-infrared hyperspectral imaging, and quantitative values of taste information (sourness and sweetness) were measured by an SA-402B electronic tongue. Considering the effects of the nonspecificity (existing in hyperspectral wavebands), interaction sensitivity, and complexity of components on the prediction of the taste information, 43 characteristic wavelengths corresponding to sourness and 22 characteristic wavelengths

corresponding to sweetness were extracted from 256 wavelengths ranging from 380 to 1038 nm by CARS, with redundant information in the hyperspectral data eliminated. In addition, the dimension of the data and the computational complexity were reduced. By comparing and analyzing the CARS–PSO–SVR model based on the characteristic wavelength and the PSO–SVR model based on all the wavelengths, it was concluded that the performance of the CARS–PSO–SVR model was superior to that of the PSO–SVR model. We showed that the redundant information caused by the complexity of the apple taste information components could be removed from the hyperspectral data using the CARS model.

For the CARS–PSO–SVR model, the R^2 values of the sourness training set and prediction set were 0.91 and 0.81, RMSEC and RMSEP were 0.017 and 0.03, respectively, and the calculation time was 39 s. On the other hand, the R^2 values of the sweetness training set and prediction set were 0.91 and 0.81, RMSEC and RMSEP were 0.042 and 0.018, respectively, and the calculation time was 78 s. The error between the actual values and the predicted values was small and the calculation time of the constructed model was short. The results showed that the taste information of apples can be obtained by the CARS–PSO–SVR model based on hyperspectral data with high accuracy, high stability, and a short calculation time, making it suitable for online nondestructive testing.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant nos. 31871882, 31772059, and 31401569), the Key Science and Technology Project of Jilin Province (20170204004SF), the State Scholarship Fund of China Scholarship Council (201808220037), China Postdoctoral Science Foundation (2018M642440), and the Project of Jilin Science and Technology Innovation and Development Plan (201751206).

References

- 1 M. Mingioni, E. Mehinagic, K. Siucińska, D. Konopacka, G. Artigas, R. Symoneaux, and I. Maitre: *Food Qual. Preference* **59** (2017) 59. <https://doi.org/10.1016/j.foodqual.2017.02.007>
- 2 A. Zdunek and J. Cybulska: *J. Sens.* **11** (2011) 6317. <https://doi.org/10.3390/s110606317>
- 3 X. Zou and J. Zhao: *Apple Quality Assessment by Fusion Three Sensors (Sensors IEEE, 2005)* 4. <https://doi.org/10.1109/icsens.2005.1597717>
- 4 É. S. Siguemoto, E. Purgatto, N. M. A. Hassimotto, and J. A. W. Gut: *LWT Food Sci. Technol.* **111** (2019) 853 <https://doi.org/10.1016/j.lwt.2019.05.111>
- 5 K. Hecke, K. Herbinger, R. Veberič, M. Trobec, H. Toplak, F. Štampar, H. Keppel, and D. Grill: *Eur. J. Clin. Nutr.* **60** (2006) 1136. <https://doi.org/10.1038/sj.ejcn.1602430>
- 6 X. Xu, Y. Bao, B. Wu, F. Lao, X. Hu, and J. Wu: *Food Chem.* **289** (2019) 250. <https://doi.org/10.1016/j.foodchem.2019.03.068>
- 7 N. Valente, A. Rudnitskaya, J. Oliveira, and E. Gaspar: *J. Sens.* **18** (2018) 2415. <https://doi.org/10.3390/s18082415>
- 8 C. Phat, B. K. Moon, and C. Lee: *Food Chem.* **192** (2016) 1068. <https://doi.org/10.1016/j.foodchem.2015.07.113>
- 9 C. Erkinbaev, K. Derksen, and J. Paliwal: *Infrared Phys. Technol.* **98** (2019) 250. <https://doi.org/10.1016/j.infrared.2019.03.033>
- 10 J. Song, M. Hu, J. Wang, M. Zhou, L. Sun, S. Qiu, Q. Li, Z. Sun, and Y. Wang: *Infrared Phys. Technol.* **96** (2019) 267. <https://doi.org/10.1016/j.infrared.2018.12.001>
- 11 L. Coic, P. Sacré, A. Dispas, A. Sakira, M. Fillet, R. D. Marini, P. Hubert, and E. Ziemons: *Talanta* **198** (2019) 457. <https://doi.org/10.1016/j.talanta.2019.02.032>

- 12 A. Adamiak, A. Zdunek, A. Kurenda, and K. Rutkowski: *J. Sens.* **12** (2012) 3215. <https://doi.org/10.3390/s120303215>
- 13 M. Shuaibu, W. S. Lee, J. Schueller, P. Gader, Y. K. Hong, and S. Kim: *Comput. Electron. Agric.* **148** (2018) 45. <https://doi.org/10.1016/j.compag.2017.09.038>
- 14 J. B. Li, X. Q. Rao, and Y.B. Ying: *Comput. Electron. Agric.* **78** (2011) 38. <https://doi.org/10.1016/j.compag.2011.05.010>
- 15 H. Lee, M. S. Kim, W.-H. Lee, and B.-K. Cho: *Sens. Actuators, B* **259** (2018) 532. <https://doi.org/10.1016/j.snb.2017.12.102>
- 16 L. Kandpal, H. Lee, M. Kim, C. Mo, and B.-K. Cho: *J. Sens.* **13** (2013) 13289. <https://doi.org/10.3390/s131013289>
- 17 A. Siedliska, P. Baranowski, and W. Mazurek: *Comput. Electron. Agric.* **106** (2014) 66. <https://doi.org/10.1016/j.compag.2014.05.012>
- 18 G. ElMasry, N. Wang, C. Vigneault, J. Qiao, and A. Elsayed: *LWT Food Sci. Technol.* **41** (2008) 337. <https://doi.org/10.1016/j.lwt.2007.02.022>
- 19 W. Tan, L. Sun, F. Yang, W. Che, D. Ye, D. Zhang, and B. Zou: *Optik* **154** (2018) 581. <https://doi.org/10.1016/j.ijleo.2017.10.090>
- 20 Q. Zhu, J. Guan, M. Huang, R. Lu, and F. Mendoza: *Postharvest Biol. Technol.* **114** (2016) 86. <https://doi.org/10.1016/j.postharvbio.2015.12.007>
- 21 M. Huang and R. Lu: *Postharvest Biol. Technol.* **58** (2010) 168. <https://doi.org/10.1016/j.postharvbio.2010.08.002>
- 22 X. Tian, J. Li, Q. Wang, S. Fan, W. Huang, and C. Zhao: *Biosyst. Eng.* **183** (2019) 110. <https://doi.org/10.1016/j.biosystemseng.2019.04.012>
- 23 Y. Ogen, C. Neumann, S. Chabrilat, N. Goldshleger, and E. B. Dor: *Geoderma* **321** (2018) 100. <https://doi.org/10.1016/j.geoderma.2018.02.011>
- 24 V. M. Gomes, A. M. Fernandes, A. Faia, and P. Melo-Pinto: *Comput. Electron. Agric.* **140** (2017) 244. <https://doi.org/10.1016/j.compag.2017.06.009>
- 25 R. Noonan: *Partial Least Squares: The Gestation Period* (Springer Cham., Berlin, 2017) pp. 3–18. https://doi.org/10.1007/978-3-319-64069-3_1

Abbreviations

Abbreviation	Full name
CARS	Competitive adaptive reweighted sampling
PSO	Particle swarm optimization
SVR	Support vector regression
RMSEP	Root mean square error of prediction set
S–G	Savitzky–Golay
R^2	Correlation coefficient
RMSEC	Root mean square error of calibration set
SPXY	Sample set partitioning based on point x - y distance
PLS	Partial least squares
ARS	Adaptive reweighted sampling
MCS	Monte Carlo sampling
EDF	Exponentially decreasing function
SVM	Support vector machine
KKT	Karush–Kuhn–Tucker conditions