

Research Article

Hyperspectral Imaging Coupled with Random Frog and Calibration Models for Assessment of Total Soluble Solids in Mulberries

Yan-Ru Zhao, Ke-Qiang Yu, and Yong He

College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China

Correspondence should be addressed to Yong He; yhe@zju.edu.cn

Received 24 April 2015; Accepted 26 August 2015

Academic Editor: Miguel de la Guardia

Copyright © 2015 Yan-Ru Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chemometrics methods coupled with hyperspectral imaging technology in visible and near infrared (Vis/NIR) region (380–1030 nm) were introduced to assess total soluble solids (TSS) in mulberries. Hyperspectral images of 310 mulberries were acquired by hyperspectral reflectance imaging system (512 bands) and their corresponding TSS contents were measured by a Brix meter. Random frog (RF) method was used to select important wavelengths from the full wavelengths. TSS values in mulberry fruits were predicted by partial least squares regression (PLSR) and least-square support vector machine (LS-SVM) models based on full wavelengths and the selected important wavelengths. The optimal PLSR model with 23 important wavelengths was employed to visualise the spatial distribution of TSS in tested samples, and TSS concentrations in mulberries were revealed through the TSS spatial distribution. The results declared that hyperspectral imaging is promising for determining the spatial distribution of TSS content in mulberry fruits, which provides a reference for detecting the internal quality of fruits.

1. Introduction

Hyperspectral imaging, as a tool for spectrochemical analysis, integrates the advantage of conventional imaging and spectroscopic technique, which can obtain both spatial and spectral information from a tested object and has been widely used in detecting quality of fruit products [1]. Hyperspectral images, known as “hypercubes,” are made up of hundreds of contiguous wavebands for each spatial position of a target. “Hypercubes” are three-dimensional ($x \times y \times \lambda$) blocks of data, comprising two spatial dimensions (x and y direction) and one wavelength dimension (λ). Hyperspectral images often contain mass wavebands, which can result in modelling complicated. In addition, the existence of multicollinearity problem could reduce the accuracy of the calibration models. Feature transformation and variables selection are being able to reduce hyperspectral data dimension. Feature transformation is a process that creates a new set of features [2] and this method has been applied to extract feature in hyperspectral data. Zou et al. [3] used principal component analysis (PCA) and independent component analysis (ICA) to reduce

the spectral dimension of the hyperspectral reflectance images of cucumber leaves and combined them with linear regression model to estimate chlorophyll concentration based on the extracted PCs and ICs. Yu et al. [4] implemented minimum noise fraction (MNF) rotation on important wavebands to extract the defective feature of hyperspectral images of loquat fruits and finally obtained that the identification accuracy was 92.3%. However, those methods only were used to eliminate useless information in view of spectra and neglect the relationship between spectral values and chemical concentrations. Feature transformation aims at preserving the topological structure of the data whereas the variables selection aims at enhancing the predictive power [2]. Recently, many studies proved that the selection of important variables can predigest calibration modelling and improve the results in terms of accuracy and robustness [5]. Moreover, selection of important wavelengths instead of full spectra of hyperspectral images has an advantage to generate chemical spatial distribution and provide a reference for developing portable multispectral imager [6]. Therefore, it is

meaningful to extract important variables to establish robust calibration model.

Random frog (RF) methodology is a novel and efficient technique for variable selection, which borrows the framework like reversible jump Markov Chain Monte Carlo (RJMCMC) [7, 8]. It executes a search in the model space through both fixed-dimensional and transdimensional moves between different models, and then a pseudo-MCMC chain is computed and used to calculate selection probability (SP) for each variable. Important variables can be selected in terms of the ranking of all variables based on SP. RF has been used to select important wavelengths from spectral data in many studies. Hu et al. [9] used a combination of RF selected reflectance and transmittance spectra from hyperspectral data to predict blueberry mechanical properties, and the results showed that prediction models based on RF had similar results with full spectral model. Li et al. [10] detected tea polyphenols (TP) of 14 cultivars of tea using infrared spectroscopy with important wavenumbers selected by interval partial least squares (iPLS) combined with RF; finally, a linear formula with 18 wavenumbers provided satisfactory results for predicting TP measurement. Yu et al. [11] employed RF and partial least squares regression (PLSR) to establish calibration model for predicting total nitrogen content of pepper plant on the basis of hyperspectral imaging in the region of 380–1030 nm.

The “hypercube” could provide visualization of biochemical constituents of a sample by calculating the chemical value of each pixel based on the spectral prediction model. There are many ways to develop calibration models, such as principal components regression (PCR), multiple linear regression (MLR), partial least squares regression (PLSR), backpropagation neural network (BPNN), and least-square support vector regression (LS-SVR). The widespread uses of PLSR make it possible to process visualization of hyperspectral images. Jin et al. [12] applied RCs-PLSR model to shift the spectrum of each pixel into its MC value for visualizing the MC distribution map in peanut kernels. K value spatial distribution map in grass carp and silver carp fillets was generated by employing the successive projection algorithm-(SPA-) PLSR model of the hyperspectral images (400–1000 nm) [13]. Simplified MLR model was used to visualize the thiobarbituric acid (TBA) values distribution in fish fillets and obtained good results (of 0.8395 and $RMSE_p$ of 0.1147 mg MDA/kg flesh) [14]. Some authors attempted to realize chemical concentration visualization of hyperspectral images using nonlinear calibration model. For example, BPNN model considerably improved the performance of prediction set (R_p^2 of 0.938 and 0.965, RPD of 4.590 and 9.335) for detecting lycopene and total phenolic content in intact tomatoes and the BPNN model made it possible to predict the bioactive compounds in each pixel of the hyperspectral images [15]. Yang et al. [16] used hyperspectral imaging technique to detect different browning levels of lychee pericarp fruits that were affected by moisture contents. A few studies reported that SPA-LS-SVM was successfully applied to generate various indexes (freshness [17], total viable counts (TVC) [18], and Warner-Bratzler shear force (WBSF) [19]) distribution of

meat for detecting quality of meat products. Other methods also could be used to map component distribution based on hyperspectral images; Siripatrawan and Makino [20] detected fungal infection on brown rice grains at early stage by using unsupervised self-organizing map (SOM) and visualized data classification of different levels of fungal infection.

Mulberry (*Fructus Mori*) fruits with bumpy surface and good taste were popular in food processing. Consumers' perception and satisfaction regarding fruit quality is an important issue in marketing [21]. Total soluble solids (TSS), which include the carbohydrates, organic acids, proteins, fats, and minerals, can contribute to the quality of fruits. Consequently, there is a need in the main producing industries to determine TSS rapidly and nondestructively to assure that fruits meet a minimum level of acceptance. Spectroscopic techniques have been proved to be an effective approach to detect the internal quality of fresh fruits [21]. Unfortunately, the spectroscopic technology fails to provide the quality parameters spatial information, which is essential to detail the analysis of the products' features [22]. What is more, based on the scanning mode, spectroscopic technology generally detects the small areas of fruits' surface and the information of the whole sample might be lost in the testing process. Visualization of chemical components using hyperspectral imaging has received an increasing attention in food-processing industry [23].

The specific objectives of this study were to (1) investigate the potential of Vis-NIR hyperspectral imaging to detect TSS of the mulberry fruits, (2) select the important wavelengths using RF algorithm, (3) evaluate the performance of the different calibration models with RF-PLSR and RF-LS-SVM based on linear and RBF kernel function, (4) compare the TSS visualization results which were developed with PLSR and LS-SVM and select the optimal calibration model, and (5) predict TSS value of each pixel in tested samples on the basis of the optimal calibration model and generate TSS spatial distribution.

2. Materials and Methods

2.1. Fruit Samples. Mulberries from local orchards (Hangzhou, Zhejiang, China) were selected for the research. The fruits were harvested randomly, ensuring a wide range of TSS. Prior to measurement, all 310 fruits were mature with the absence of any green area. Each single fruit constituted a sample, and as a result, 310 samples were collected and stored in the refrigerator at a constant temperature at 3°C in the laboratory before the hyperspectral images were acquired. Samples were removed from the refrigerator and placed under room condition (~20°C) for more than 2 hours.

2.2. Hyperspectral Image Acquisition and Calibration. Mulberry samples were scanned by a push-broom hyperspectral imaging apparatus with reflectance mode as shown in Figure 1. The hyperspectral imaging system mainly consisted of an imaging spectrograph (ImSpectorV10E, Spectral Imaging Ltd., Finland) covering the spectral range of 380–1,030 nm; a CCD camera (C8484-05, Hamamatsu, Japan) coupled with a zoom lens (OLES23, Specim, Spectral Imaging

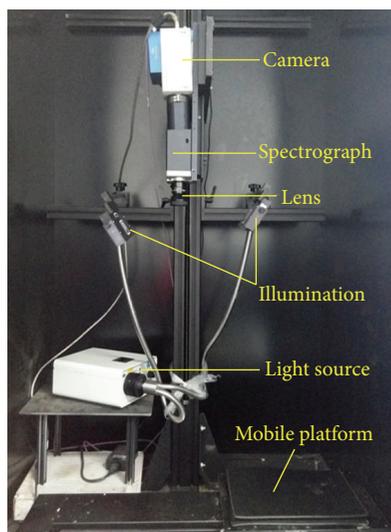


FIGURE 1: Schematic diagram of the main components of the hyperspectral imaging.

Ltd., Oulu, Finland); an assembled illumination source coupled with two 150-W quartz tungsten halogen lamps (Fiber-Lite DC950 Illuminator, Dolan Jenner Industries Inc., USA); a mobile platform operated by a stepper motor (IRCP0076, Isuzu Optics Corp., Taiwan); and a computer with the spectral imaging system V10E software (Isuzu Optics Corp., Taiwan), which was used to set and adjust the parameters of the device, including exposure time, motor speed, imaging acquisition, wavelength range, and image correction. The spectral resolution is 2.8 nm; the resolution of CCD camera is 672×512 (spatial \times spectral) pixels. Some parameters of apparatus for acquiring hyperspectral images should be set and adjusted before acquiring hyperspectral images of mulberries. In this work, the moving speed of mobile platform is 1.6 mm/s, exposure time of the CCD camera is 0.008 s, and the distance from the lens to samples is 295 mm. The whole system (except the computer) was assembled in a dark chamber to minimize the effects of ambient light during the sample scanning [4, 24].

Due to the existence of dark current in CCD camera and the uneven intensity of the light source in different bands, several bands with weaker light intensity contained the biggest noises [24]. Here, raw hyperspectral images were calibrated using the white and dark reference based on (1) for weakening the effect of dark current in the CCD camera and the uneven intensity of light in different bands:

$$I = \frac{I_{\text{raw}} - I_{\text{dark}}}{I_{\text{white}} - I_{\text{dark}}}, \quad (1)$$

where I is the calibrated hyperspectral images, I_{raw} is the raw hyperspectral images, I_{white} is the white reference images ($\sim 99\%$ reflectance), and I_{dark} is the dark reference images ($\sim 0\%$ reflectance).

2.3. Image Processing. Hyperspectral data were extracted from the calibrated hyperspectral images using ENVI software (version 4.6, ITT Visual Information Solutions, Boulder, USA). Before acquiring accurate spectra of the samples, background information should be removed in batches and this process has shorter processing time than manually extracted ROI of the individual sample. “Imsubtract” and “Threshold” algorithm were employed to create the mask. The details are displayed in Figure 2. The results subtracted from waveband at 893 nm (a) and 569 nm (b) are shown in Figure 2(c), in which it was found that there were big differences on gray value between tested samples and background images. Then threshold of 0.4 was used to remove the background and good masks (d) were obtained.

By implementing the masks in the original hyperspectral images, the background information was removed. The separated region was identified as the region of interest (ROI) of the sample. The average spectrum of spectra of all pixels in a mulberry was considered as the spectrum of a sample and 310 spectrums were collected. To avoid the low signal-noise ratio and diminish the problem of high dimensionality of feature spaces, the wavebands of 420–1,000 nm were considered in the analysis [11]. Ultimately, a spectral data matrix of 310×460 (samples \times wavebands) was obtained for further analysis.

2.4. Detection of TSS. Total soluble solids (TSS, °Brix%) in mulberry samples were measured using traditional destructive tests. After acquiring the hyperspectral data, each fruit unit was juiced and TSS were measured by using WAY-2S Digital Refractometer (Shanghai Precision & Scientific Instrument Co., Ltd., Shanghai, China). The instrument range covered from 0 to 95% with temperature correction and refractive index accuracy is ± 0.0002 [25]. All measurements were averaged over the data from three replicates in a room at 20°C.

2.5. Chemometrics Methods

2.5.1. Random Frog Algorithm. The key steps of RF are illustrated in Figure 3 and the detailed algorithm of RF was described in literatures [7, 8]. Before running RF algorithm, five parameters (T , Q , θ , ω , and η) should be assigned to proper values. T was the number of iterations and needed to be sufficiently large to achieve convergence ($T = 10,000$); Q was the number of variables in the initialized variables set ($Q = 50$); θ controlled variance of a normal distribution ($\theta = 0.3$); ω was a coefficient explained in Step 2 ($\omega = 3$); η represented the upper bound of the probability ($\eta = 0.1$) [7].

2.5.2. Partial Least Squares Regression. Partial least squares regression (PLSR) is a multivariate data analysis technique which generalizes and combines features from principal component analysis (PCA) and multiple linear regression (MLR) [26]. PLSR has been successfully used in developing multivariate calibration models, as it uses the concentration information (\mathbf{Y}) in determining how regression factors are computed from the spectral data matrix (\mathbf{X}), thereby reducing the impact of irrelevant X -variations in the calibration

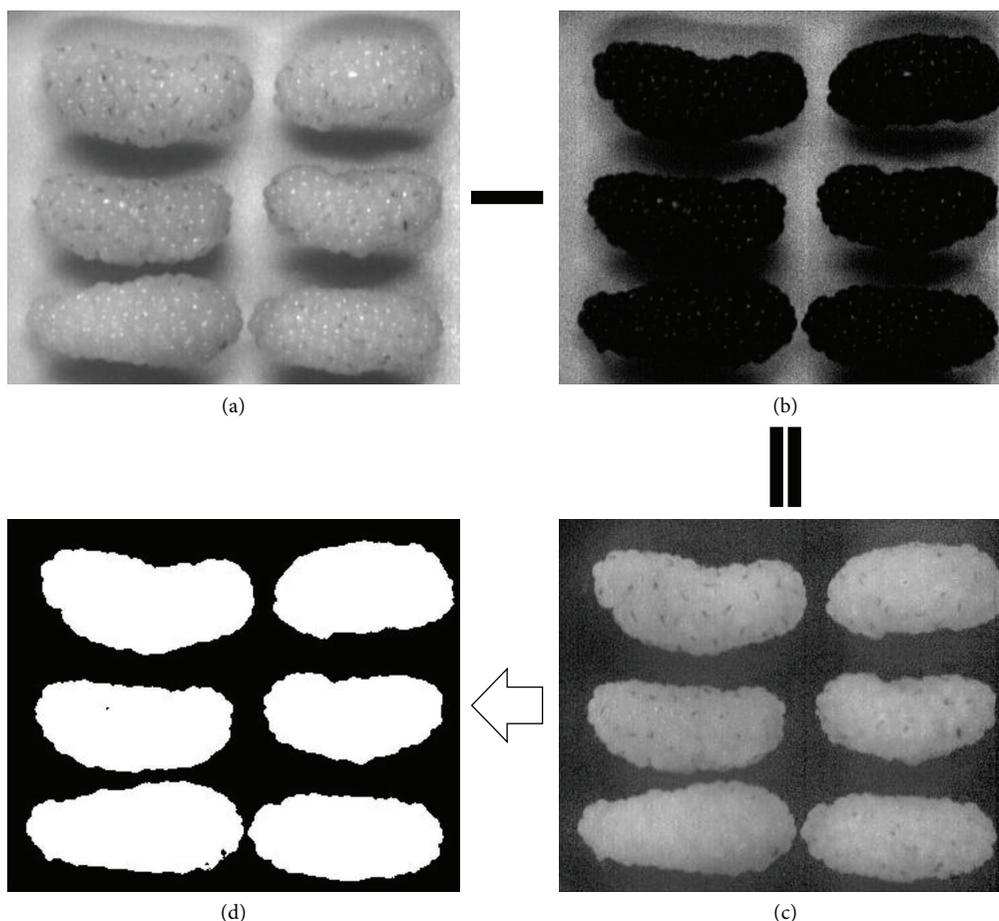


FIGURE 2: Details of acquiring mask (a) gray image at 893 nm; (b) gray image at 569 nm; (c) insubtract result of (a) and (b); (d) mask image.

model. The important feature of PLSR is based on latent variables (LVs) [27].

There were three steps to develop PLSR; the first step is to decompose the matrix and the model is

$$\begin{aligned} \mathbf{X} &= \mathbf{TP} + \mathbf{E}, \\ \mathbf{Y} &= \mathbf{UQ} + \mathbf{F}, \end{aligned} \quad (2)$$

where \mathbf{T} and \mathbf{U} are the score matrices of \mathbf{X} matrix and \mathbf{Y} matrix, \mathbf{P} and \mathbf{Q} are the loading matrices of \mathbf{X} matrix and \mathbf{Y} matrix, and \mathbf{E} and \mathbf{F} are the errors which come from the process of PLS.

The second step is to process \mathbf{T} and \mathbf{U} by linear regression. It must build the following linear correlation:

$$\mathbf{U} = \mathbf{BT} + \mathbf{E}, \quad (3)$$

where $\mathbf{B} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{Y}$.

Finally, unknown \mathbf{Y} is predicted in the following PLSR model:

$$\mathbf{Y}_p = \mathbf{T}_p \mathbf{BQ}. \quad (4)$$

2.5.3. Least-Square Support Vector Machine. LS-SVM is an alternate formulation of SVM regression proposed by

Suykens et al. [28]. The main advantage is that it is computationally more efficient than the standard SVM method. The details of LS-SVM algorithm were introduced as follows [5].

Optimization problem of LS-SVM is formulated:

$$\min J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2 \quad (5)$$

subject to the constraints

$$y_k = w^T \varphi(x_k) + b + e_k, \quad k = 1, \dots, N, \quad (6)$$

where γ is the regularization parameter which balances the model's complexity and the training errors; e_k is the random error; x_k and y_k are input and output variables; k is sample number.

And then, Lagrange function is applied to solve the optimization problem

$$\begin{aligned} L(w, b, e, \alpha) &= J(w, e) \\ &- \sum_{k=1}^N \alpha_k \{w^T \varphi(x_k) + b + e_k - y_k\}, \end{aligned} \quad (7)$$

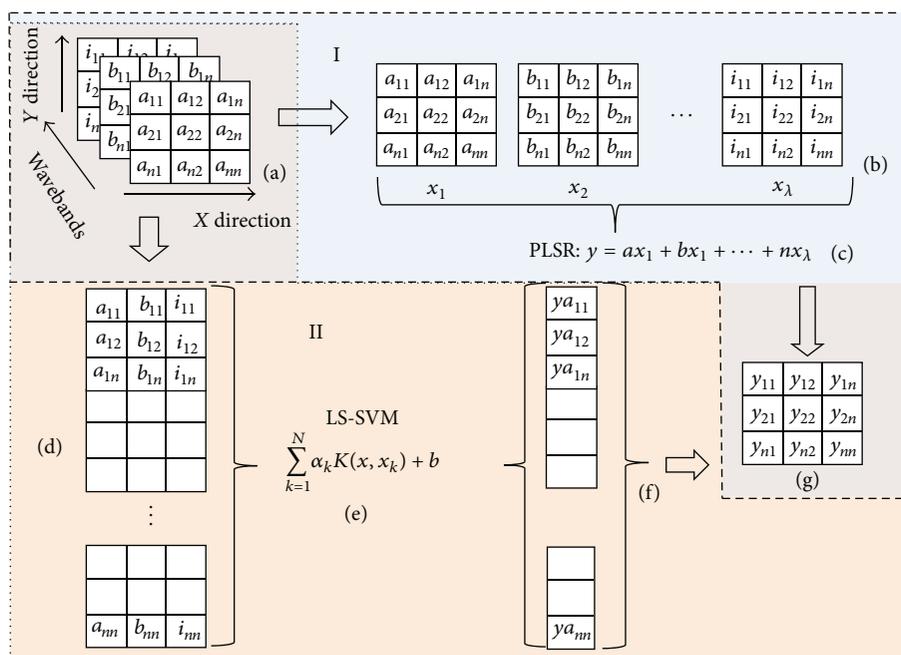


FIGURE 4: Visualization flowchart based on PLSR and LS-SVM.

of pixels, $1 \times n^2$ data (e) was figured out, and then (e) was folded into a $n \times n$ matrix (f), namely, the image of the tested sample with predictive TSS value. Pseudo-colour images were created with different colours representing different levels of TSS that were predicted by the optimal simplified model [32]. Useful information about TSS distribution in mulberries was observed by checking chemical images.

There are many kinds of noise in hyperspectral images, such as electrical noise from CCD detector; the noise was caused by transmitting procedure and others. The presence of the noise seriously affects the feature extraction and recognition accuracy of the tested objects. It is necessary to process image denoising before conducting further analysis. Median filtering, in which gray value of every pixel is set to be the average gray value in a certain neighbourhood window, is a nonlinear filtering technique that has been successfully applied to many signal and image processing tasks [33]. Median filtering (5×5) was applied to denoise hyperspectral images in this study.

2.8. Software. Random frog (version 2.0) toolbox, LS-SVM lab v1.8 toolbox, median filtering toolbox, and the visualization program were finished with MATLAB R2009a (version 7.8) software. PLSR models were established using “The Unscrambler X10.1” (CAMO PROCESS AS, Oslo, Norway). Origin Pro 8.0 SR0 (Origin Lab Corporation, Northampton, MA, USA) software was employed to design the curve figures. All the processes were run on a PC (CPU: G2020 @ 2.90 GHz, RAM 4.00 GB) operating on Windows 7.

3. Results and Discussion

3.1. Analysis of the Measured TSS. In order to establish the calibration models of the TSS in mulberries, all mulberry

TABLE 1: Summary of the measured TSS in samples of calibration and prediction set ($^{\circ}$ Brix).

Group	Range	Mean	SD ^a
Calibration	10.99–3.21	6.86	1.56
Prediction	9.86–3.86	6.35	1.45
Total	10.99–3.21	6.75	1.55

^aStandard deviation of each group.

samples were divided into a calibration set and prediction set according to the ratio about 3 : 1 (235 : 75) by SPXY method, in which sample set partitioning is based on joint x - y distances [34]. Table 1 summarized the measured TSS of all samples, the maximum and minimum value of calibration set were 10.99° Brix and 3.21° Brix, and the range of prediction set was 9.86 – 3.86° Brix. Usually, calibration set with a wide range of the chemical component could establish a robust model.

3.2. Feature of Spectra. Spectra of all mulberries covering the range of 420–1,000 nm are displayed in Figure 5. Lower reflectance ($<10\%$) in the visible region of 420–650 nm was attributed to the relatively homogeneous and dark purple or black colour of mature fruits and mainly caused by anthocyanin and chlorophyll [35, 36]. Reflectance for mulberry samples started to increase dramatically from 650 to 800 nm and reached a peak at 850 nm. An obvious valley located around 960–980 nm, which was attributed to the combination effect of -OH groups from carbohydrate and water [37]. This fact has also been reported in the case of Huang et al. [25]. Chemometrics methods were introduced to analyze the spectra and establish the relationship between spectra and measured TSS to determine the internal quality of mulberries in the future study.

TABLE 2: Results of TSS predictive models of mulberries.

Number	Models	No. ^c	LVs ^d	Calibration		Cross-validation		Prediction		$\gamma(10^3)$	$\sigma^2(10^3)$
				R_C	RMSE _C	R_{CV}	RMSE _{CV}	R_P	RMSE _P		
(1)	Full-PLSR	460	21	0.984	0.279	0.958	0.445	0.959	0.411		
(2)	Full-LS-SVM-RBF	460		0.999	<0.1	0.999	<0.1	0.999	<0.1	4.43×10^5	
(3)	Full-LS-SVM-linear	460		0.999	<0.1	0.999	<0.1	0.999	<0.1	1.11×10^8	8.99×10^2
(4)	RF-PLSR	23	13	0.942	0.522	0.928	0.579	0.899	0.675		
(5)	RF-LS-SVM-linear	23		0.944	0.515	0.931	0.571	0.886	0.714	4.93	
(6)	RF-LS-SVM-RBF	23		0.999	0.061	0.958	0.453	0.956	0.430	9.56	4.29
(7)	RF-PLSR	11	7	0.811	0.912	0.792	0.951	0.834	0.800		
(8)	RF-LS-SVM-linear	11		0.818	0.899	0.796	0.945	0.843	0.781	0.36	
(9)	RF-LS-SVM-RBF	11		0.984	0.283	0.912	0.647	0.925	0.557	1.25×10^8	1.80

^cNumbers of the wavelengths used for analysis.

^dLatent variables of the PLSR model.

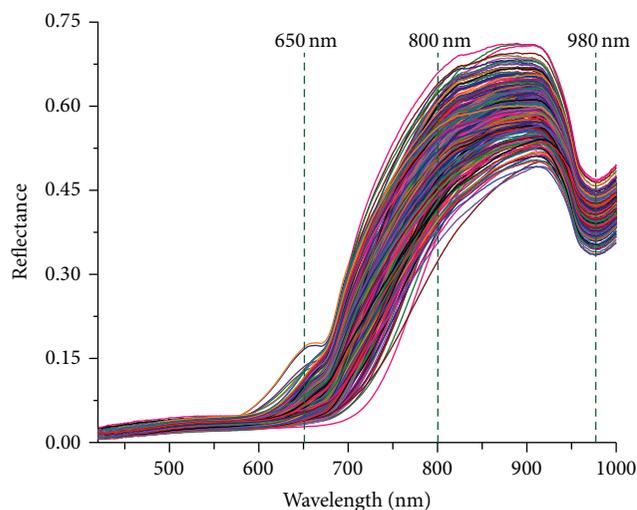


FIGURE 5: Spectral curves of all mulberries covering the range of 420–1,000 nm.

3.3. Selection of the Important Wavelengths. RF was used to select important wavelengths from original data. Figure 6 displays SPs of wavelengths, and a small number of wavelengths displayed had high SP (over 0.9); most of the wavelengths were with low SP, and these results showed that there existed a lot of useless information to make TSS content be predicted using hyperspectral imaging. If the cutoff of SP was 0.7 (1) and 0.85 (2), there were 23 (960, 929, 814, 849, 432, 500, 877, 995, 903, 915, 831, 956, 967, 954, 923, 511, 454, 840, 835, 527, 509, 850, and 883 nm) and 11 (960, 929, 814, 849, 432, 500, 877, 995, 903, 915, and 831 nm) important wavelengths selected, respectively. These wavebands were ranked by descending order of SPs. These wavelengths were mainly concentrated on two parts: 400–550 and 800–1,000 nm. The concentration of anthocyanin and chlorophyll caused low reference in the visible region and contributed to the maturity of mulberry fruits. Wavelengths at 814 nm, 975–989 nm, and 981 nm correspond to second and third overtones of $-\text{NH}$, 877 nm and 900–950 nm were assigned to $-\text{CH}$ third overtone, and 960 nm was assigned to

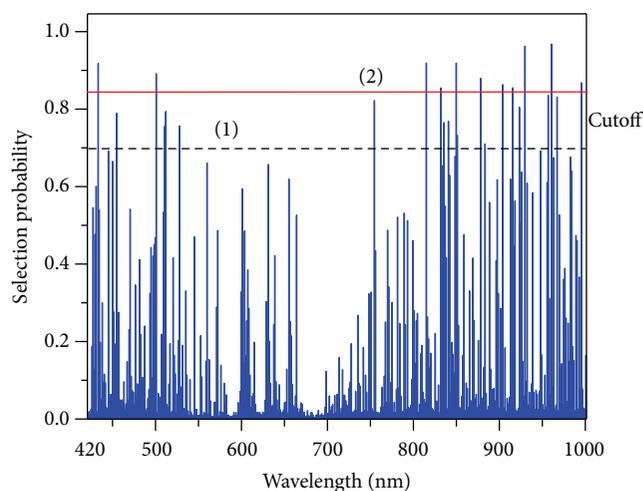


FIGURE 6: Selection probability of each wavelength.

$-\text{OH}$ second overtone. These bands are assigned to mono- and poly-carbohydrates (fructose, glucose, and pectin) and water in mulberry [36].

In order to establish calibration models with fewer features, two important wavelengths sets (23 versus 11) were employed to establish calibration models and these results were compared to obtain optimal visualization map of TSS in mulberry.

3.4. Calibration Models. Multivariate analyses, developed with leave-one-out CV, were used to find accurate PLSR and LS-SVM models for the prediction of TSS. The predictive models of TSS in mulberry fruits were built using the two kinds of the selected wavelengths (23 versus 11) and the results of these models are enumerated in Table 2.

Overall, LS-SVM regression models had better performance for predicting TSS than PLSR models because LS-SVM is a nonlinear regression model and it could transform the original data into a high dimension space to make linear solution [38]. LS-SVM was capable of solving the nonlinear problem of the calibration models. LS-SVM with RBF

kernel function based on 23 wavelengths with R_p of 0.956 and $RMSE_p$ of 0.430 could provide the most effective TSS estimation compared to other models, while LS-SVM models with linear function had similar results with RF-PLSR models (except with full wavelengths). This consequence caused by the RBF kernel function of LS-SVM has an advantage in conducting samples in multidimensional space. In addition, linear kernel function in LS-SVM was considered as a special form of RBF kernel function [39].

LS-SVM with fewer wavelengths was able to provide more accurate results. Model (9) with 11 important wavelengths had an approving expression with R_p of 0.925 and $RMSE_p$ of 0.557, and these results were equal to the performance of the full-PLSR model (R_p of 0.959 and $RMSE_p$ of 0.411). When the number of important wavelengths reduced to almost a half (11 versus 23), model (9) only had a little reduction of 3.24% (R_p) and an increase of 29.5% ($RMSE_p$) compared to model (6). Hence, model (9) is appropriate to predict TSS in mulberry fruits.

Although LS-SVM had better performance, PLSR with 23 important wavelengths offered acceptable results. Model (4) provided a reliable result of R_p of 0.899 and $RMSE_p$ of 0.675. Compared with full-PLSR model, R_C , R_{CV} , and R_p in model (4) showed a slight reduction of 3.88%, 2.52%, and 5.47%; $RMSE_C$, $RMSE_{CV}$, and $RMSE_p$ provided a change of 0.213, 0.101, and 0.228, respectively. However, 95.0% of the variables (23 versus 460) were removed in RF-PLSR model. RF-PLSR model could provide a more effective prediction for TSS of mulberries. The accuracy of RF-PLSR model was higher than SPA-MLR model in the literature of [25] for predicting TSS of mulberries, because hyperspectral imaging could provide both spectra and image information about tested samples. In addition, spectra of whole fruit were averaged as the spectrum of the sample to avoid the loss of spectra. But beyond that, there were only two wavelengths in the SPA-MLR model which meant that useful spectral information might be over-removed. However, RF-PLSR model for predicting TSS of mulberries was not as good as Monte Carlo-uninformative variable elimination- (MC-UVE-) SPA (MC-UVE-SPA) model for predicting TSS of "Ya" pear [31]. This phenomenon could be explained that the mulberry fruit has a bumpy surface that has more influence on the spectra than does a smooth surface [25]. TSS prediction of mulberries based on RF-PLSR model was parallel to the prediction TSS of blueberries that were acquired by interval partial least square- (iPLS-) PLSR model [35].

3.5. Distribution Map of TSS in Mulberries. In order to seek an optimal calibration model for realizing TSS visualization in mulberry fruits, models (4), (8), and (9) were employed to process visualization procedure. The comparative maps of six samples developed by three models are exhibited in Figure 7.

Six mulberry fruits were used to compare the reliability of TSS distribution maps, which were predicted by three models. PLSR (Figure 7(a)) and LS-SVM with linear kernel function (Figure 7(b)) could provide clearly TSS distribution, while LS-SVM with RBF kernel function (Figure 7(c)) failed to display TSS visualization of mulberry fruits. There are

four possible reasons to explain this phenomenon: (1) the special modelling way of LS-SVM with RBF kernel function, which needed to transfer the raw data into high-dimensional space, might change the original data form; (2) there were two parameters (γ, σ^2) to control the predicted results of LS-SVM, which might add the complexity of LS-SVM; (3) using calibration model with 235 variables to predict a map (such a map of a mulberry was about 100×200 pixels) might produce problem of overfitting; (4) the mulberry fruit with bumpy appearance would bring the variation of spectral reflectance, which might affect the accuracy of models.

The linear kernel function is the special form of RBF kernel function, and LS-SVM with linear kernel function (LS-SVM-linear) (b) had better TSS distribution (b) than that of LS-SVM-RBF. However, in Figure 7(b), it was hard to distinguish the TSS level of mulberry and a mistake was made in sample (2). The measured TSS value of sample (2) was 8.2°Brix and it was the highest TSS value among these six samples, while in Figure 7(b), it had the lightest colour. PLSR had the best performance compared to LS-SVM for its simple linear combination. At the same time, the fitting effects of PLSR and LS-SVM with linear kernel function (LS-SVM-linear) were compared and expressed in Figure 7(d). The correlation coefficients between measured TSS values and predicted TSS values in PLSR and LS-SVM with linear kernel function (LS-SVM-linear) were 0.857 and -0.500 , respectively. Although LS-SVM could offer satisfactory results in calibration model, PLSR with simple algorithm was feasible to map TSS distribution in mulberry fruits. When the simplified model was finally established, it was subsequently employed to predict TSS in each pixel of the image resulting in new pseudo-colour images and this process was called "prediction map" [32]. As the last step of analyzing hyperspectral image, RF-PLSR model was used to predict TSS of each pixel and transferred its hyperspectral image to the TSS distribution map.

The multilinear function for the TSS prediction of the mulberries was obtained:

$$\begin{aligned}
 Y_{TSS} = & 16.207 - 0.191\lambda_1 - 0.818\lambda_2 - 1.717\lambda_3 \\
 & - 1.481\lambda_4 + 0.007\lambda_5 + 0.037\lambda_6 - 1.291\lambda_7 \\
 & - 0.193\lambda_8 - 1.112\lambda_9 - 1.017\lambda_{10} - 1.606\lambda_{11} \\
 & - 0.230\lambda_{12} - 0.168\lambda_{13} - 0.132\lambda_{14} - 0.924\lambda_{15} \quad (12) \\
 & + 0.042\lambda_{16} + 0.023\lambda_{17} - 1.541\lambda_{18} - 1.574\lambda_{19} \\
 & + 0.047\lambda_{20} + 0.040\lambda_{21} - 1.474\lambda_{22} \\
 & - 1.252\lambda_{23},
 \end{aligned}$$

where λ_i is the spectral reflectance value at the wavelength of i nm (i was 960, 929, 814, 849, 432, 500, 877, 995, 903, 915, 831, 956, 967, 954, 923, 511, 454, 840, 835, 527, 509, 850, and 883, resp.) and Y_{TSS} is the predictive TSS values of the mulberry fruits.

New hyperspectral data with only 23 wavelengths could speed up the visualization process and make it easier to establish a multispectral imaging system. The function (12)

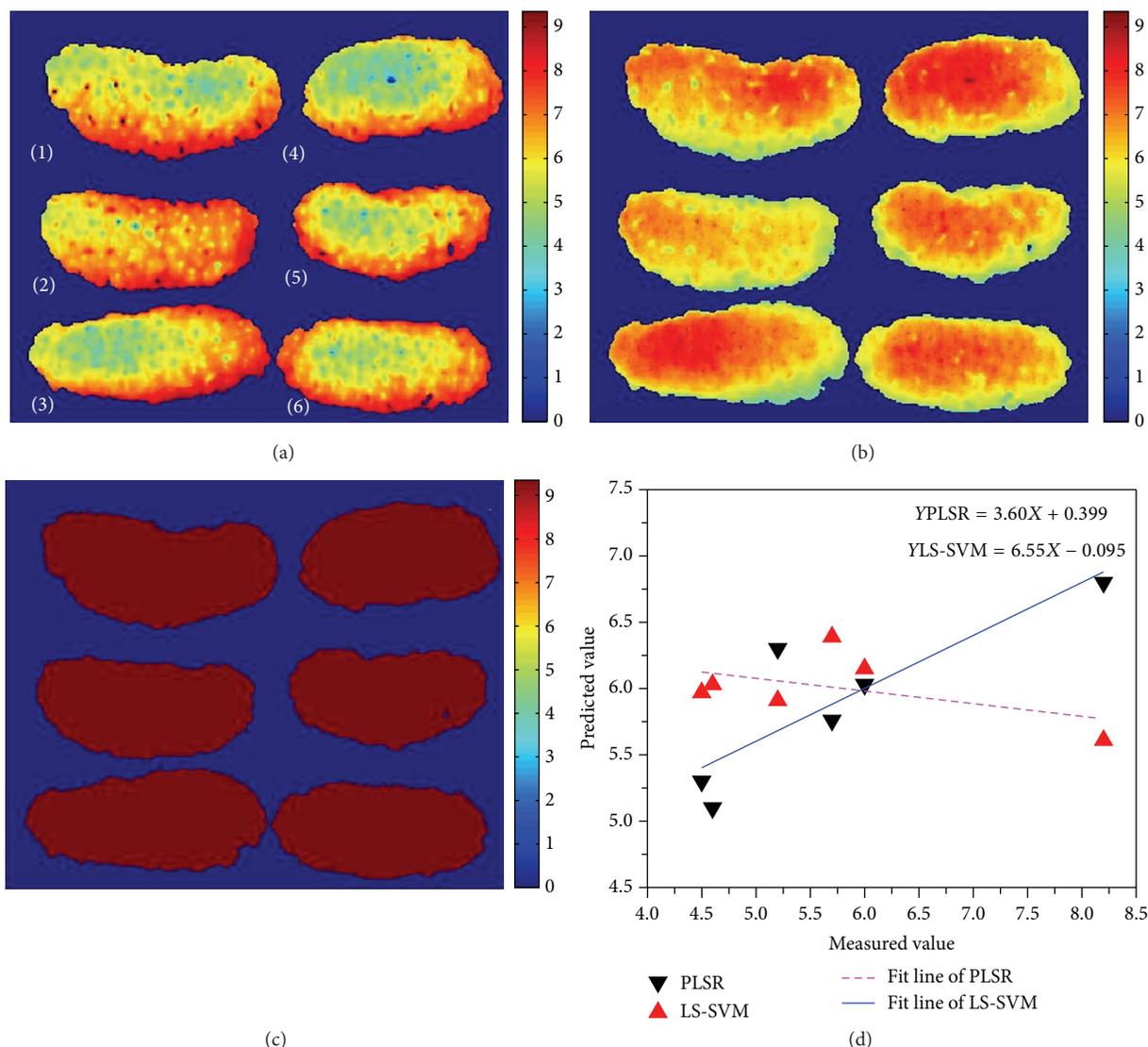


FIGURE 7: Compared TSS distribution results with PLSR and LS-SVM models: (a) distribution map using PLSR model; (b) visualization result based on LS-SVM model with linear kernel function; (c) TSS distribution using LS-SVM with RBF kernel function; (d) fitting effect between measured and predicted results of PLSR and LS-SVM with linear kernel function. Colour bar represents the TSS concentration.

was employed to predict the TSS of each pixel within the mulberry fruits hyperspectral images. The spatial distribution maps of samples' TSS are generated in Figure 8.

The predicted TSS concentration of each pixel was mapped with a linear colour scale using different colours from red to blue to represent different TSS concentrations from high to low. Mulberries with higher TSS values have more pixels shown in red, such as in Figure 8 (1–6) samples. The average TSS value of these six samples was 9.3°Brix. The more the pixels coloured in green and blue, the lower the TSS values. There were unsatisfying results: the blue pixels in samples (13) and (14) were saturated points and were not considered in the analysis [40].

Different values in the maps of TSS distribution were in quantitative proportion to the spectrum of the corresponding pixels. However, these distribution maps are difficult to be

inspected by naked eyes [41]. The ability to provide spatial information makes hyperspectral imaging available to focus on detecting both external and internal quality of fruits.

4. Conclusions

To sum up, the successful mapping of TSS distribution in mulberries suggested that the application of hyperspectral imaging to realize the visualization of mulberry fruits' internal quality is feasible and promising. The PLSR and LS-SVM model based on 23 and 11 wavelengths had a good performance to predict TSS of mulberries, which indicated that RF algorithm was effective in reducing three-dimensional data. PLSR-RF based on 23 important wavelengths provided the optimal visualization results. It could be revealed that PLSR was feasible to map chemical component concentration (TSS)

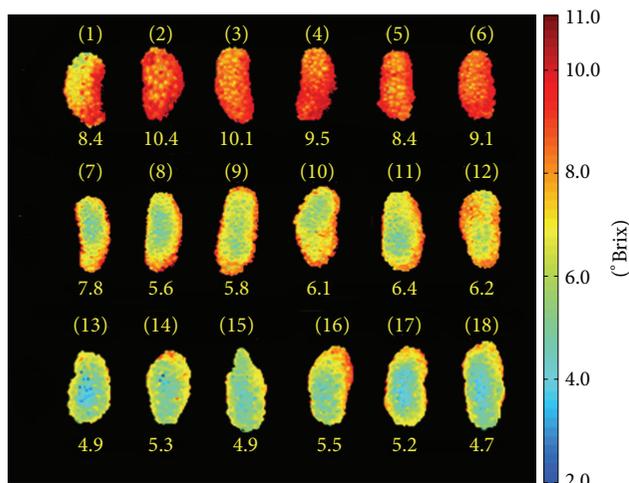


FIGURE 8: TSS visualization of mulberry fruits.

distribution of mulberry fruits. This research provided a theoretical basis for developing the instrument for measuring the internal quality of fruits and made it possible to sort mulberries based on TSS spatial distribution.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This study was supported by 863 National High-Tech Research and Development Plan (Project no. 2013AA102301) and the Fundamental Research Funds for the Central Universities.

References

- [1] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, and J. M. Frias, "Hyperspectral imaging—an emerging process analytical tool for food quality and safety control," *Trends in Food Science and Technology*, vol. 18, no. 12, pp. 590–598, 2007.
- [2] H. Liu and H. Motoda, "Feature transformation and subset selection," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 2, pp. 26–28, 1998.
- [3] X. Zou, J. Zhao, M. Holmes et al., "Independent component analysis in information extraction from visible/near-infrared hyperspectral imaging data of cucumber leaves," *Chemometrics and Intelligent Laboratory Systems*, vol. 104, no. 2, pp. 265–270, 2010.
- [4] K.-Q. Yu, Y.-R. Zhao, Z.-Y. Liu, X.-L. Li, F. Liu, and Y. He, "Application of visible and near-infrared hyperspectral imaging for detection of defective features in loquat," *Food and Bioprocess Technology*, vol. 7, no. 11, pp. 3077–3087, 2014.
- [5] D. Wu, Y. He, S. Feng, and D.-W. Sun, "Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM," *Journal of Food Engineering*, vol. 84, no. 1, pp. 124–131, 2008.
- [6] M. Kamruzzaman, G. Elmasry, D.-W. Sun, and P. Allen, "Application of NIR hyperspectral imaging for discrimination of lamb muscles," *Journal of Food Engineering*, vol. 104, no. 3, pp. 332–340, 2011.
- [7] H.-D. Li, Q.-S. Xu, and Y.-Z. Liang, "Random frog: an efficient reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification," *Analytica Chimica Acta*, vol. 740, pp. 20–26, 2012.
- [8] Y.-H. Yun, H.-D. Li, L. R. E. Wood et al., "An efficient method of wavelength interval selection based on random frog for multivariate spectral calibration," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 111, pp. 31–36, 2013.
- [9] M. Hu, Q. Dong, B. Liu, U. L. Opara, and L. Chen, "Estimating blueberry mechanical properties based on random frog selected hyperspectral data," *Postharvest Biology and Technology*, vol. 106, pp. 1–10, 2015.
- [10] X. Li, C. Sun, L. Luo, and Y. He, "Determination of tea polyphenols content by infrared spectroscopy coupled with iPLS and random frog techniques," *Computers and Electronics in Agriculture*, vol. 112, pp. 28–35, 2015.
- [11] K.-Q. Yu, Y.-R. Zhao, X.-L. Li, Y.-N. Shao, F. Liu, and Y. He, "Hyperspectral imaging for mapping of total nitrogen spatial distribution in pepper plant," *PLoS ONE*, vol. 9, no. 12, Article ID e116205, 2014.
- [12] H. Jin, L. Li, and J. Cheng, "Rapid and non-destructive determination of moisture content of peanut kernels using hyperspectral imaging technique," *Food Analytical Methods*, 2015.
- [13] J. Cheng, D. Sun, H. Pu, and Z. Zhu, "Development of hyperspectral imaging coupled with chemometric analysis to monitor K value for evaluation of chemical spoilage in fish fillets," *Food Chemistry*, vol. 185, pp. 245–253, 2015.
- [14] J.-H. Cheng, D.-W. Sun, H.-B. Pu, Q.-J. Wang, and Y.-N. Chen, "Suitability of hyperspectral imaging for rapid evaluation of thiobarbituric acid (TBA) value in grass carp (*Ctenopharyngodon idella*) fillet," *Food Chemistry*, vol. 171, pp. 258–265, 2015.
- [15] C. Liu, W. Liu, W. Chen, J. Yang, and L. Zheng, "Feasibility in multispectral imaging for predicting the content of bioactive compounds in intact tomato fruit," *Food Chemistry*, vol. 173, pp. 482–488, 2015.
- [16] Y.-C. Yang, D.-W. Sun, and N.-N. Wang, "Rapid detection of browning levels of lychee pericarp as affected by moisture contents using hyperspectral imaging," *Computers and Electronics in Agriculture*, vol. 113, pp. 203–212, 2015.
- [17] Q. Dai, J.-H. Cheng, D.-W. Sun, H. Pu, X.-A. Zeng, and Z. Xiong, "Potential of visible/near-infrared hyperspectral imaging for rapid detection of freshness in unfrozen and frozen prawns," *Journal of Food Engineering*, vol. 149, pp. 97–104, 2015.
- [18] J. H. Cheng and D. W. Sun, "Rapid and non-invasive detection of fish microbial spoilage by visible and near infrared hyperspectral imaging and multivariate analysis," *LWT—Food Science and Technology*, vol. 62, no. 2, pp. 1060–1068, 2015.
- [19] H.-J. He, D. Wu, and D.-W. Sun, "Potential of hyperspectral imaging combined with chemometric analysis for assessing and visualising tenderness distribution in raw farmed salmon fillets," *Journal of Food Engineering*, vol. 126, pp. 156–164, 2014.
- [20] U. Siripatrawan and Y. Makino, "Monitoring fungal growth on brown rice grains using rapid and non-destructive hyperspectral imaging," *International Journal of Food Microbiology*, vol. 199, pp. 93–100, 2015.
- [21] J. A. Cayuela, "Vis/NIR soluble solids prediction in intact oranges (*Citrus sinensis* L.) cv. Valencia Late by reflectance,"

- Postharvest Biology and Technology*, vol. 47, no. 1, pp. 75–80, 2008.
- [22] F. Mendoza, R. Lu, D. Ariana, H. Cen, and B. Bailey, “Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content,” *Postharvest Biology and Technology*, vol. 62, no. 2, pp. 149–160, 2011.
- [23] G. ElMasry, M. Kamruzzaman, D.-W. Sun, and P. Allen, “Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review,” *Critical Reviews in Food Science and Nutrition*, vol. 52, no. 11, pp. 999–1023, 2012.
- [24] K. Yu, Y. Zhao, X. Li, Y. Shao, F. Zhu, and Y. He, “Identification of crack features in fresh jujube using Vis/NIR hyperspectral imaging combined with image processing,” *Computers and Electronics in Agriculture*, vol. 103, pp. 1–10, 2014.
- [25] L. Huang, D. Wu, H. Jin, J. Zhang, Y. He, and C. Lou, “Internal quality determination of fruit with bumpy surface using visible and near infrared spectroscopy and chemometrics: a case study with mulberry fruit,” *Biosystems Engineering*, vol. 109, no. 4, pp. 377–384, 2011.
- [26] N. Shetty and R. Gislum, “Quantification of fructan concentration in grasses using NIR spectroscopy and PLSR,” *Field Crops Research*, vol. 120, no. 1, pp. 31–37, 2011.
- [27] I. S. Helland, “Some theoretical aspects of partial least squares regression,” *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 97–107, 2001.
- [28] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, “Weighted least squares support vector machines: robustness and sparse approximation,” *Neurocomputing*, vol. 48, pp. 85–105, 2002.
- [29] C. Wang, “The nuclear techniques and the selection of model parameters in big data,” *International Journal of Database Theory and Application*, vol. 7, pp. 151–162, 2014.
- [30] Z. Hu, Y. Bao, and T. Xiong, “Electricity load forecasting using support vector regression with memetic algorithms,” *The Scientific World Journal*, vol. 2013, Article ID 292575, 10 pages, 2013.
- [31] J. Li, W. Huang, L. Chen et al., “Variable selection in visible and near-infrared spectral analysis for noninvasive determination of soluble solids content of ‘Ya’ pear,” *Food Analytical Methods*, vol. 7, no. 9, pp. 1891–1902, 2014.
- [32] Y.-Z. Feng, G. Elmasry, D.-W. Sun, A. G. M. Scannell, D. Walsh, and N. Morcy, “Near-infrared hyperspectral imaging and partial least squares regression for rapid and reagentless determination of Enterobacteriaceae on chicken fillets,” *Food Chemistry*, vol. 138, no. 2-3, pp. 1829–1836, 2013.
- [33] A. C. Bovik, T. S. Huang, and D. C. Munson, “The effect of median filtering on edge estimation and detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 2, pp. 181–194, 1987.
- [34] R. K. H. Galvão, M. C. U. Araujo, G. E. José, M. J. C. Pontes, E. C. Silva, and T. C. B. Saldanha, “A method for calibration and validation subset partitioning,” *Talanta*, vol. 67, no. 4, pp. 736–740, 2005.
- [35] G. A. Leiva-Valenzuela, R. Lu, and J. M. Aguilera, “Assessment of internal quality of blueberries using hyperspectral transmittance and reflectance images with whole spectra or selected wavelengths,” *Innovative Food Science and Emerging Technologies*, vol. 24, pp. 2–13, 2014.
- [36] B. N. Peshlov, F. E. Dowelt, F. A. Drummond, and D. W. Donahue, “Comparison of three near infrared spectrophotometers for infestation detection in wild blueberries using multivariate calibration models,” *Journal of Near Infrared Spectroscopy*, vol. 17, no. 4, pp. 203–212, 2009.
- [37] G. A. Leiva-Valenzuela, R. Lu, and J. M. Aguilera, “Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging,” *Journal of Food Engineering*, vol. 115, no. 1, pp. 91–98, 2013.
- [38] J. Valyon and G. Horváth, “A robust LS-SVM regression,” *Proceedings of World Academy of Science, Engineering and Technology*, vol. 7, pp. 148–153, 2005.
- [39] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, “An introduction to kernel-based learning algorithms,” *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–201, 2001.
- [40] Z. Schmilovitch, T. Ignat, V. Alchanatis, J. Gatker, V. Ostrovsky, and J. Felföldi, “Hyperspectral imaging of intact bell peppers,” *Biosystems Engineering*, vol. 117, pp. 83–93, 2014.
- [41] D. Wu, H. Shi, Y. He, X. Yu, and Y. Bao, “Potential of hyperspectral imaging and multivariate analysis for rapid and non-invasive detection of gelatin adulteration in prawn,” *Journal of Food Engineering*, vol. 119, no. 3, pp. 680–686, 2013.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

