Food Chemistry 218 (2017) 413–421

Contents lists available at ScienceDirect

Food Chemistry

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

Analytical Methods

How to predict the sugariness and hardness of melons: A near-infrared hyperspectral imaging method

Meijun Sun [a], Dong Zhang [a], Li Liu [b], Zheng Wang [c,*]

[a] School of Computer Science and Technology, Tianjin University, Tianjin 300072, China
[b] School of Environmental Science and Engineering, Tianjin University, Tianjin 300072, China
[c] School of Computer Software, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 18 March 2016
Received in revised form 18 July 2016
Accepted 5 September 2016
Available online 7 September 2016

Keywords:
Hyperspectral image
Melon
Non-intrusive quality measurement
Sweetness
Hardness

ABSTRACT

Hyperspectral imaging (HSI) in the near-infrared (NIR) region (900–1700 nm) was used for non-intrusive quality measurements (of sweetness and texture) in melons. First, HSI data from melon samples were acquired to extract the spectral signatures. The corresponding sample sweetness and hardness values were recorded using traditional intrusive methods. Partial least squares regression (PLSR), principal component analysis (PCA), support vector machine (SVM), and artificial neural network (ANN) models were created to predict melon sweetness and hardness values from the hyperspectral data. Experimental results for the three types of melons show that PLSR produces the most accurate results. To reduce the high dimensionality of the hyperspectral data, the weighted regression coefficients of the resulting PLSR models were used to identify the most important wavelengths. On the basis of these wavelengths, each image pixel was used to visualize the sweetness and hardness in all the portions of each sample.

1. Introduction

Over the past few decades, the fruit and vegetable industry has grown in terms of both international and domestic markets (Desmond, Kenny, Ward, & Sun, 2000; McDonald, Sun, & Kenny, 2001). Consumers prefer to purchase high-quality products for the best possible price. To satisfy these demands, it is essential to provide an objective method for measuring the quality of fruits and vegetables. Healthy appearance and sweetness are the two primary quality-related metrics that are of concern to consumers (Herrero, 2008).

The melon (*Cucumis melo* L.) was one of the earliest domesticated crops. It is widely grown because of its high economic value. The majority of consumers appreciate the melon for its nutrients, unique flavour and aroma. In addition, the melon is one of the top ten fruits on the international market (Menon Soumya & Rao, 2014). With the more stringent market for fruit quality, the requirements for melons are no longer limited to fruit traits such as beautiful appearance, good quality and long shelf life. Instead, a pleasant aroma and a crisp taste have become important factors for attracting consumers and for enhancing market competitiveness. Unfortunately, current techniques focus more on the colour, size, disease resistance, yield and other traits of the fruit, and there are few considerations relating to fruit firmness and crisp taste. Therefore, this deficiency must be addressed with quality measurements because consumers pay more attention to the sweetness of melons.

Traditional methods for assessing fruit quality are destructive and time consuming. Improvements in the control and classification of mechanized processes and instrumental techniques for rapidly screening fruit properties are of significant interest for both the industry and the consumers (Gowen, O'Donnell, & Cullen, 2007; Lorente et al., 2012). However, these systems have limitations that need to be augmented with other suitable systems to measure the desired quality traits (Pu, Feng, & Sun, 2015). Machine vision-supported food analysis has increased considerably in recent years. This approach has been applied to meat, fish, grains, bread, fruits and vegetables (Mery, Pedreschi, & Soto, 2015; Pu, Sun, & Ma, 2015; Zhang, Huang, & Li, 2014). Machine vision techniques can provide substantial information about the nature and attributes of objects. In this paper, we use another important machine vision feature that opens up the electromagnetic spectrum, which is inaccessible to the human eye (Chagnot et al., 2015; Jiang, Qiao, & He, 2016; Ma, Sun, & Pu, 2016). In this study, we use a near-infrared (NIR) hyperspectral imaging system.

Among the techniques proposed for evaluating food quality, near-infrared spectroscopy has been shown to have significant potential for the continuous monitoring and control of food

* Corresponding author.
E-mail address: wzheng@tju.edu.cn (Z. Wang).

http://dx.doi.org/10.1016/j.foodchem.2016.09.023
0308-8146/

industry processes (Bock & Connelly, 2008; Cen & He, 2007). First, the NIR hyperspectral imaging system is fast and can evaluate melon quality within a shorter time than can conventional monitoring. Second, NIR HSI technology provides complete information on the spectral constituents for a large range of scanned samples. Most importantly, NIR HSI technology is non-intrusive and non-destructive, which is essential in fruit quality measurement. In contrast to traditional methods, the NIR HSI technique is fast, simple and able to obtain spectral information in a non-intrusive way. Thus, it can potentially replace expensive and slow reference methods (Heigl, Petter, Lieb, Bonn, & Huck, 2009; Prevolnik, Candek-Potokar, & Skorjanc, 2010). Spectral techniques in the visible and near-infrared ranges have already been widely applied to food, including meat products. Common applications include bruise detection in white mushrooms, crystal pears and oranges, surface damage detection in apples, sour skin damage detection in onions, pit detection in cherries, and quality attribute evaluations of strawberries (ElMasry, Wang, ElSayed, & Ngadi, 2007; Peng & Lu, 2006; Qin, Burks, Zhao, Niphadkar, & Ritenour, 2012; Wang, Li, Tollner, Gitaitis, & Rains, 2012). However, there are few applications that employ hyperspectral imaging systems for predicting melon quality.

Because hyperspectral imaging is an extension of both spectroscopy and imaging techniques, it integrates both spectral and spatial backgrounds in one system (He & Sun, 2015; Theron et al., 2015). In recent years, researchers in the academic and industrial sectors have become increasingly interested in non-destructive analysis. Hyperspectral imaging provides abundant spectral information. Hyperspectral imaging is used to identify the types of material because different materials show unique spectral curves. Therefore, the potential use of hyperspectral imaging for quality assessments of fruits on processing lines is very promising. Because the dimensionality of each pixel on a hyperspectral image is very high, so a dimensionality reduction is necessary for classification and computing efficiency. We have improved this technique by seeking the most important bands for predicting the essential quality attributes that contribute to the development of the optimized "multispectral" imaging system; these improvements can be implemented directly for industrial applications.

2. Materials and methods

2.1. Preparation of the melon samples

There were 60 samples from three types of melons (Elizabeth, Claw Brittle, and Emerald) in this study. In total, 20 samples were picked from each melon species from a vegetable cultivation area in Tianjin, China. The melons were cut in half along the longitudinal direction. Hyperspectral images of the samples were then obtained. Afterwards, the samples were analyzed at the measurement centre to measure their sweetness and hardness parameters.

2.2. Measurements of sweetness and hardness

After all the melons were prepared, we needed to acquire their hyperspectral images, and reference values for sweetness and hardness. Melon sweetness was measured with a Brix detector (REFRACTOMETER PAL-1). Sweetness values were obtained from eight different locations on each sample half. The first two locations were on the seam: one was close to the flesh, and the other was close to the peel. The third and fourth locations were on the melon tail: one was close to the flesh and the other was close to the peel. The last four locations were on the fruit equator.

Hardness values were tested for each sample, using a texture detector. The same locations were used for calculating hardness values as for sweetness measurements. However, we only selected four locations to calculate hardness. The values of each attribute for all the tested melon samples were then used as reference values; these values were to calibrate the system to predict the given attributes. In Table 1, the relevant statistics for the sweetness and hardness of three types of melon samples are presented.

2.3. Hyperspectral imaging system

In order to obtain the hyperspectral images of melons, a near-infrared hyperspectral imaging acquisition system was set up first. As shown in Fig. 1, the system consisted of a CCD hyperspectral camera, a translation stage, illumination units and a reflector. The CCD hyperspectral camera used was a line-scanning type (also called 'Pushbroom'), it records a line of a hyperspectral image rather than the whole space at a time. A narrow line of the melon is imaged onto a row of pixels in the hyperspectral image and the spectrograph generates a spectrum for each point on the line (ElMasry, Sun, & Allen, 2012). Due to the line scanning, a translation stage was needed to move the entire range of the melon. An 800 W halogen lamp was used as the illumination unit in order to provide enough light, and reduce the external interference. The camera resolution is 320×400 pixels. All the image acquisition parameters (i.e., the translation stage motor speed and exposure time of the camera) were controlled with computer software.The near-infrared spectral range (882–1719 nm) was used in this study. The spectral resolution was approximately 3 nm. Thus, a total of 256 spectral bands were produced by the hyperspectral image acquisition system. Under the same experimental conditions, we could obtain NIR hyperspectral images that provide both spectral and spatial information by constantly changing the melon samples. We refer to these images as the raw image data (R_0). To reduce the impact of light and noise, dark and white reference images must also be obtained under the same conditions. The calibrated hyperspectral images can then be obtained with the following equation:

$$R(\lambda) = \frac{R_0(\lambda) - B_d}{B_w - B_d} \times 100\% \quad (1)$$

where R is the calibrated hyperspectral image data and B_d and B_w are the dark and white reference images, respectively. A standard white reference image was obtained by acquiring a spectral image from a high-reflectance (approximately 99.99%) white calibration tile, and the dark current noise image (approximately 0%) was acquired by recording a spectral image when the light was turned off and the camera lens was completely covered with a black cap. For the resulting hyperspectral images, we used ENVI software to view the spatial information in a single band. We determined that the smaller and longer wavelength band images contained considerable noise. To reduce the effects caused by the instrument itself, we used the middle 216 bands (ElMasry et al., 2012).

2.4. Spectral analysis method

2.4.1. Spectral information extraction

As mentioned in Section 2.2, the measured values for each half-melon sample were extracted at eight sweetness measurement points and four hardness measurement points. Then we scanned those half-melons to obtain a hyperspectral image for each melon and selected the regions of interest (ROI) at every exactly the same sampling location from all hyperspectral images. Each ROI contains 50–80 pixels which are averaged to calculated the spectral value of the ROI.

Table 1
Reference values of sweetness and hardness of three varieties of melon samples as collected by traditional methods.

Statistics	Elizabeth		Claw Brittle		Emerald	
	Sweetness	Hardness	Sweetness	Hardness	Sweetness	Hardness
Mean	11.3294	1851.9	6.4788	1250.8	11.8725	1245.7
Standard deviation	6.3064	33,095	3.3509	122,500	3.0056	218,900
Minimum	16.20	3221.9	11.30	2219.6	15.20	2242.2
Maximum	5.60	718.60	3.5	623.9280	6.00	235.65
Range	10.60	2503.3	7.8	1595.672	9.20	2006.55
n	160	80	160	80	160	80

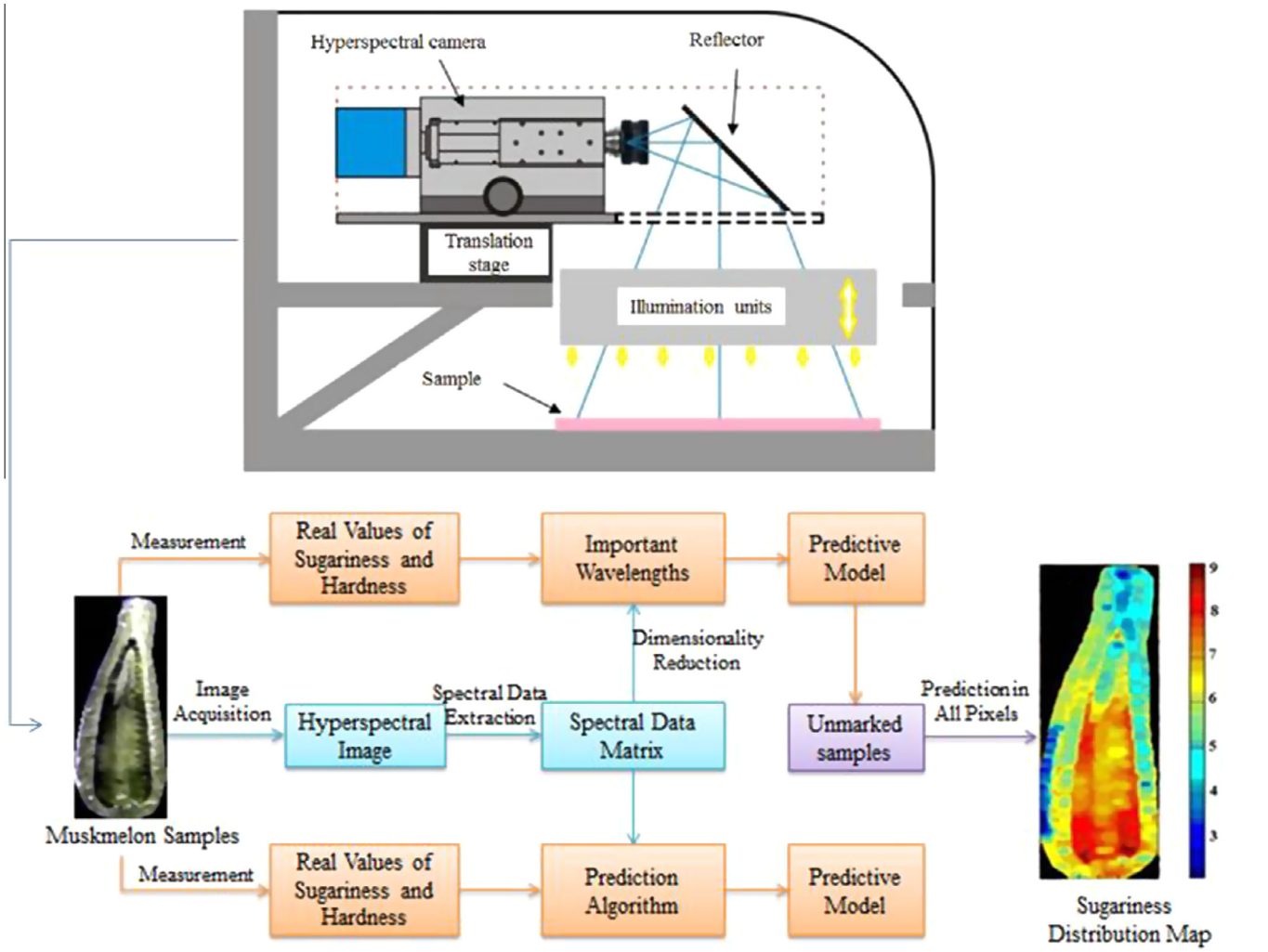

Fig. 1. Hyperspectral imaging system and flowchart of the algorithm.

2.4.2. Prediction model analysis

Classical prediction models were developed to predict melon sweetness and hardness with NIR spectral information. In this paper, partial least squares regressions (PLSRs) (Antonucci et al., 2011; Cheng & Sun, 2015; Sinija & Mishra, 2011), support vector machines (SVMs) (Camps-Valls, Gomez-Chova, & Munoz-Mari, 2006; Foody & Mathur, 2006), and artificial neural networks (ANNs) (Lorente, Aleixos, Gmez-Sanchis, Cubero, & Blasco, 2013) were used to predict these parameters. In addition, a principal component analysis was applied to reduce the data dimensionality before SVM and ANN were used.

PLSR is a quantitative decomposition technique that is used to optimize the covariance between the label and linear combinations of feature by simultaneously decomposing the input data. PLSR was used to create linear predictive models based on the spectral data (the X matrix) and the measured sweetness or hardness (the Y matrix). PCA was performed on the data reduction and it was based on identifying the most important directions of variability in a multivariate data space (the X matrix) to identify the primary phenomena in the data set. SVM was to find the optimal hyperplane to classify the dataset and it is effective for small sample size and low dimension. ANN was to imitate the human neural network to complete the prediction and it has a strong nonlinear fitting ability, and can map an arbitrarily complex nonlinear relationship.

Spectral information data were denoted as the *X* matrix, and the melon sweetness or hardness values that were measured according to a traditional measurement approach were denoted as the *Y*

matrix. The PLSR was used as the linear model for predicting X and Y. The PLSR model optimized the covariance between Y and X by simultaneously performing their decomposition.

PCA + SVM and PCA + ANN were then used as prediction models between the X and Y matrices. A principal component analysis (PCA) was performed on all the spectral data (X matrix) to identify the spectral outliers. A PCA was used to identify the most important directions of the variability in a multivariate data space (X matrix) and to determine the primary phenomena in the dataset (Cariou, Chehdi, & Le-Moan, 2011; Hou, Song, Min, & Hoon Park, 2012; Kuo & Landgrebe, 2004; Liao, Pižurica, Scheunders, Philips, & Pi, 2013). SVM and ANN used this compressed spectral data as the input and then predicted the values of the parameters (Nogalesbueno, Ayala, & Heredia, 2015).

To choose the most reliable and efficient prediction model from the above-mentioned three methods, some evaluation values were used. In general, the highest coefficient of determination (R^2) and the smallest root mean square error ($RMSE$) of the predicted values, that were calculated by using the model and the measured values, resulted in the best performance. These statistical parameters are defined as follows:

$$R^2 = \frac{\sum (y_{pred} - \overline{y_{act}})^2}{\sum (y_{act} - \overline{y_{act}})^2} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum (y_{pred} - y_{act})^2}{n}} \tag{3}$$

where n is the number of samples, y_{act} is the measured value, and y_{pred} is the predicted sweetness or hardness value that is estimated by the predictive models. All computational and multivariate data analyses were performed with Matlab.

2.4.3. Selection of important wavelengths

Spectral dimension of each pixel on the hyperspectral image is a 216-dimensional vector which is much higher than the pixel on the RGB 3-channel image. However, among the 216 vectors, there are a lot of redundant data because of the similarity between neighbour data (Osborne, Jordan, & Kunnemeyer, 1997). Thus, we had to perform a dimension reduction to speed up the analysis process and to reduce the calculation time. In most reliable applications of hyperspectral imaging, a subset of wavelengths with the most valuable and useful information can remove noise and result in lower prediction error (Indahl & Naes, 2004; Keskin, Dodd, Han, & Khalilian, 2004). In this paper, the PLSR generates the best predictive performance. Thus, we selected the important wavelengths on the basis of the PLSR model to predict the values.

In the PLSR model, an interpretation of beta-coefficients was used to determine the relevant wavelength regions. An important beta-coefficient value is considered as a relevant reflectivity region. The spectral bands that correspond to the maximum and minimum values of beta-coefficients can indicate a significant variable, and they are regarded as important wavelengths (Cariou et al., 2011; ElMasry et al., 2012; Wang et al., 2012). In this paper, the PLSR model was first used on the full spectral space and we selected important wavelengths on the basis of the beta-coefficient values. The model was then used in the important spectral space.

2.4.4. Distribution maps of melon quality attributes

After the important wavelengths were selected for the new reduced spectral space, they were used for making predictions in less time and at lower calculation costs than the traditional method. Furthermore, it was possible for us to examine all the hyperspectral images to observe the differences in quality attributes from sample to sample and even within the same sample. The prediction values were obtained by predicting all the pixels in the examined hyperspectral images and then visualizing them with a different colour to form the quality distribution images. In this study, the PLSR model exhibited the best results. Thus, we used PLSR to predict the parameters of the examined hyperspectral images. The resulting images were referred to as the "distribution maps". All of the steps that are involved in building these maps are depicted in the flowchart shown in Fig. 1.

3. Results and discussion

3.1. Spectral characteristics

After pre-processing, the hyperspectral images of the melon samples consisted of 216 wavelengths from sub-images. The hyperspectral images could also be regarded as $I(x, y, \lambda)$, in which x and y describe the spatial information of the sub-images and λ describes the spectrum information in each pixel. Any given pixel has 216 dimensions around the spectrum, and these dimensions form a spectrum line that acts as a fingerprint of physicochemical features at that position. Because a traditional measurement was used to confirm the results, we needed to select the same position as the region of interest on the hyperspectral image samples to obtain the spectral space.

In this paper, the typical spectral signatures of the samples originated from various cultivars, including Elizabeth melon, Claw Brittle melon and Emerald melon. These three melon types have different sweetness or hardness values, which are measured at different locations, as shown in Fig. 2. The sweetness and hardness are shown in Fig. 3. In Fig. 3(a), the samples share the same spectral patterns over the entire wavelength region (900–1600 nm), although there are differences for some locations. This finding is expected because different anatomical locations have different major chemical compositions, such as the compositions of sucrose and water. The Claw Brittle sample presented the highest reflectivity intensities throughout the entire spectral region. However, the Elizabeth sample had the lowest reflectivity compared with the other two.

Two classes were identified in the samples by using the sweetness and hardness values. The melon samples with high sweetness (sweetness > 10) values showed higher reflectivity intensities than did those with low sweetness values (sweetness < 10). This observation was especially pronounced in the spectral region from 1150 to 1250, as depicted in Fig. 3(b). Moreover, the reflectivity of spectral lines that indicate low hardness values (hardness < 1000) is greater than it is for high hardness values (hardness > 1000), as illustrated in Fig. 3(c). The reflectivity differences between different samples and the different parameter values in the NIR are primarily related to differences in scattering properties. The physical structure and unfixed scatter of the surface lead to various spectral motifs (ElMasry et al., 2012).

In general, different chemical compositions result in different reflectivities. The reflectance at approximately 1050 nm is clearly high for these particular ranges. However, at 1050 nm, the reflectance begins to decrease sharply, up to 1150 nm. The most prominent reflectance bands in the NIR region are related to C—H and O—H functional groups. These groups are common in major constituents (e.g., water, sucrose and cellulose) within the samples (Garini, Young, & McNamara, 2006). Additionally, the bands in the 1300–1400 nm regions are ascribed to the combination bands of C—H vibrations. Broad bands in the 1400–1600 nm regions are attributed to the overlapping first overtones of the O—H stretching modes of self-associated and water-bonded O—H functional groups. Strictly speaking, the local reflectance trough that appears

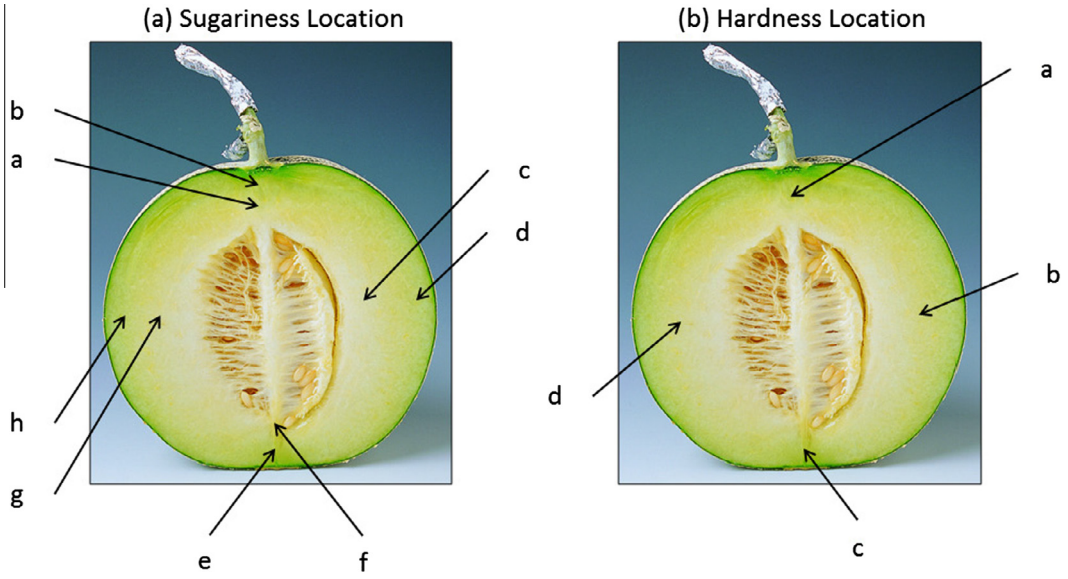

Fig. 2. (a) Eight locations from which sweetness information was extracted. (b) Four locations from which hardness information was extracted.

at approximately 1200 nm is related to the presence of water in the sample. The reflectance peak at 1060 and 1270 nm (C—H stretching second overtone) is caused by sucrose and cellulose.

3.2. Predicting sweetness and hardness by using the full spectral range

3.2.1. Prediction of sweetness

In this experiment, we calculated the R^2 and RMSE ten times and used the average value to evaluate the performance of the three models. As described in Section 2.4.2, R^2 and RMSE were calculated to measure the performance of the three models. The higher R^2 or the smaller RMSE represent the better performance of the prediction model. In the prediction process, we randomly selected the training data and testing data, and made sure that the ratio was 7:1. Five cross validation was used to get the variables: for example, the correct number of latent factors in the PLSR model, the optimal reduced dimensions and parameters of SVM in the SVM or ANN model.

The prediction of melon quality attributes was performed with the PLSR, PCA + SVM, and PCA + ANN. In this prediction step, the full spectral range of 216 wavelengths was used. Table 2 shows the primary statistics that were used to evaluate the performance of the models for predicting sweetness and hardness.

As shown in Table 2, it is clear that the PLSR models for sweetness are considerably better than the PCA + SVM and PCA + ANN. In the PLSR model, we observed that the coefficient of determination between the sample spectra and sweetness were as high as 0.7952 for the Elizabeth variety, 0.6568 for the Claw Brittle variety, and 0.5723 for the Emerald variety with root mean square errors of 1.1698, 1.1731 and 1.1771, respectively. In addition, the PLSR models that employed the NIR spectra appeared to be robust at predicting melon sweetness because a small number of latent factors was used to develop these models.

The PLSR-predicted results for the sweetness of Elizabeth samples are presented in the scatter plots shown in Fig. 4(a). In all the Figures, the ordinate and abscissa axes represent the predicted and measured fitted values of the corresponding parameters, respectively. The correlation between these values for each characteristic and the R-squared number and RMSE were embedded inside each graph to indicate the prediction performance of the models.

3.2.2. Hardness prediction

In the test model, the hardness correlation was 0.3671 for the Elizabeth variety, 0.3417 for the Claw Brittle variety, and 0.3412 for the Emerald variety, and the RMS values were 503.2680, 351.4339, and 399.6786, respectively. A scatter plot for the measured hardness values of Elizabeth in comparison with the predicted values from the NIR spectra, using the above-mentioned models, is shown in Fig. 4(b). As shown, the predicted results are not satisfactory. This poor result occurs because the primary hardness factor is associated with the cell structure and the cellulose content of the cell wall. The hyperspectral image is not sensitive to the cell structure. A different cell structure produces different hardness but has the same material composition. The cell structure, especially the cell wall, significantly affects cell hardness. Thus, when the hyperspectral system was used to predict hardness, no significant correlation was observed.

3.3. Prediction of sweetness and hardness using important wavelengths

There is a high degree of spectral redundancy within the hyperspectral image data. Irrelevant information can cause instability in the prediction models. Therefore, the important wavelengths that contain fewer bands but have the most effective information must be selected to perform predictions that obtain a more stable predictive performance. Furthermore, using important wavelengths reduces the time and calculation costs.

When using all the wavelengths, the PLSR model achieves the best result. Thus, we only selected the important wavelengths for the prediction via the PLSR model. Important wavelengths reflect the characteristics of spectra for predicting sweetness. These wavelengths are obtained on the basis of beta-coefficients (ElMasry et al., 2012). The wavelengths with a large beta-coefficient play an important role in the regression model. The high positive or negative peaks indicate that the wavelengths at these points contain important information about the quality attribute under analysis. Wavelengths with a small coefficient are negligible and make little contribution to the model's productivity. The important wavelengths for Elizabeth were selected as shown in Fig. 5 (top row). The same was for Claw Brittle and Emerald. Fifteen wavelengths (967.32, 1010.72, 1191.94, 1342.77, 1565.68 nm for Elizabeth, 989.59, 1046.26, 1128.60, 1201.05, 1335.69 nm for Claw

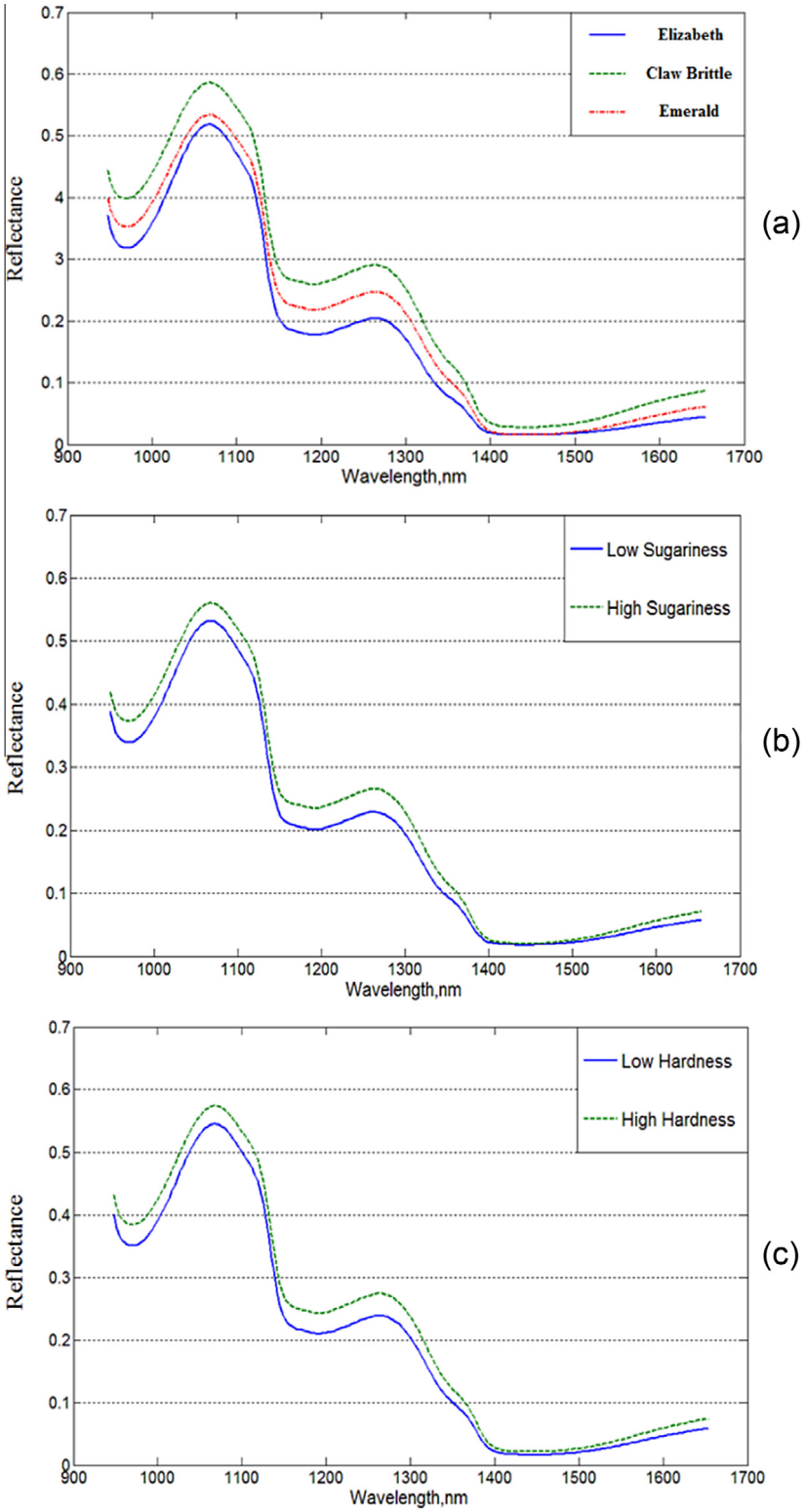

Fig. 3. Differences in the spectral profiles of the melon samples. Spectral profiles of different samples for (a) the anatomical location of the sample, (b) sweetness, and (c) hardness.

Brittle, 1036.85, 1092.68, 1141.71, 1263.75, 1335.69 nm for Emerald) were identified as being important for predicting sweetness. Additionally, the following wavelengths (990.52, 1208.24, 1421.48, 1478.77, 1619.82 nm for Elizabeth, 1035.61, 1173.25, 1235.14, 13981.41, 1501.56 nm for Claw Brittle, and 1063.75, 1315.14, 1434.27, 1492.60, 1585.12 nm for Emerald) were determined to be important for predicting hardness.

The sensitive wavelengths were selected by using the previous analysis from the entire spectrum of wavelengths to form a reduced spectral space. These spectral data were then used as the input for the PLSR model to predict the melon quality. The selection of the sensitive wavelengths is shown in Table 3 and Fig. 5. The results of the new spectral space for the Elizabeth variety in terms of R-squared were 0.7755 and 0.3555, with RMSE

Table 2
Three models for predicting sweetness and hardness that use the entire spectral range (216 wavelengths).

	PLSR		PCA + SVM		PCA + ANN	
	R-squared	RMSE	R-squared	RMSE	R-squared	RMSE
Elizabeth sweetness	0.80	1.17	0.43	2.10	0.78	1.20
Elizabeth hardness	0.37	503.23	0.20	553.47	0.37	492.72
Claw Brittle sweetness	0.66	1.17	0.39	1.49	0.60	1.21
Claw Brittle hardness	0.34	351.43	0.19	377.23	0.31	415.12
Emerald sweetness	0.57	1.18	0.27	1.47	0.51	1.22
Emerald hardness	0.34	399.68	0.32	414.05	0.34	397.20

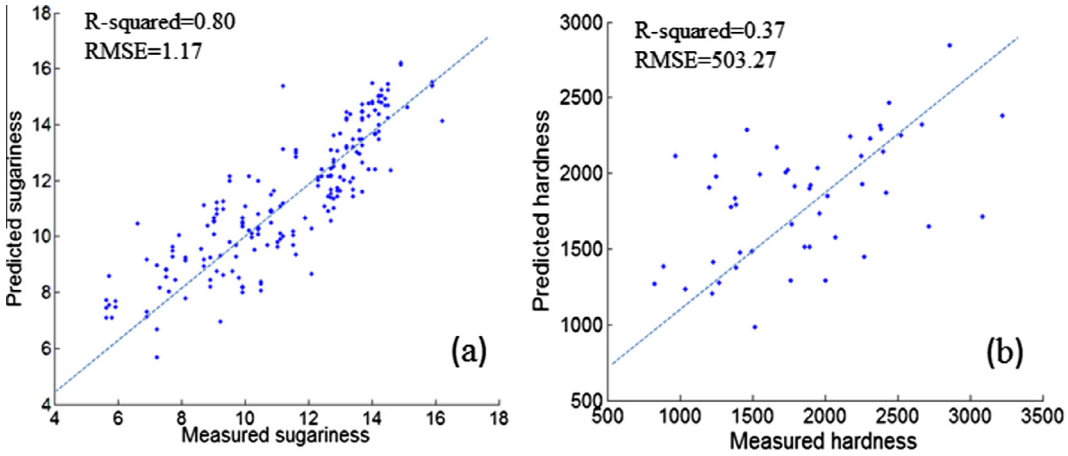

Fig. 4. Comparison of the measured and predicted values for Elizabeth samples by using the PLSR model with the entire spectral range. (a) Sweetness and (b) hardness.

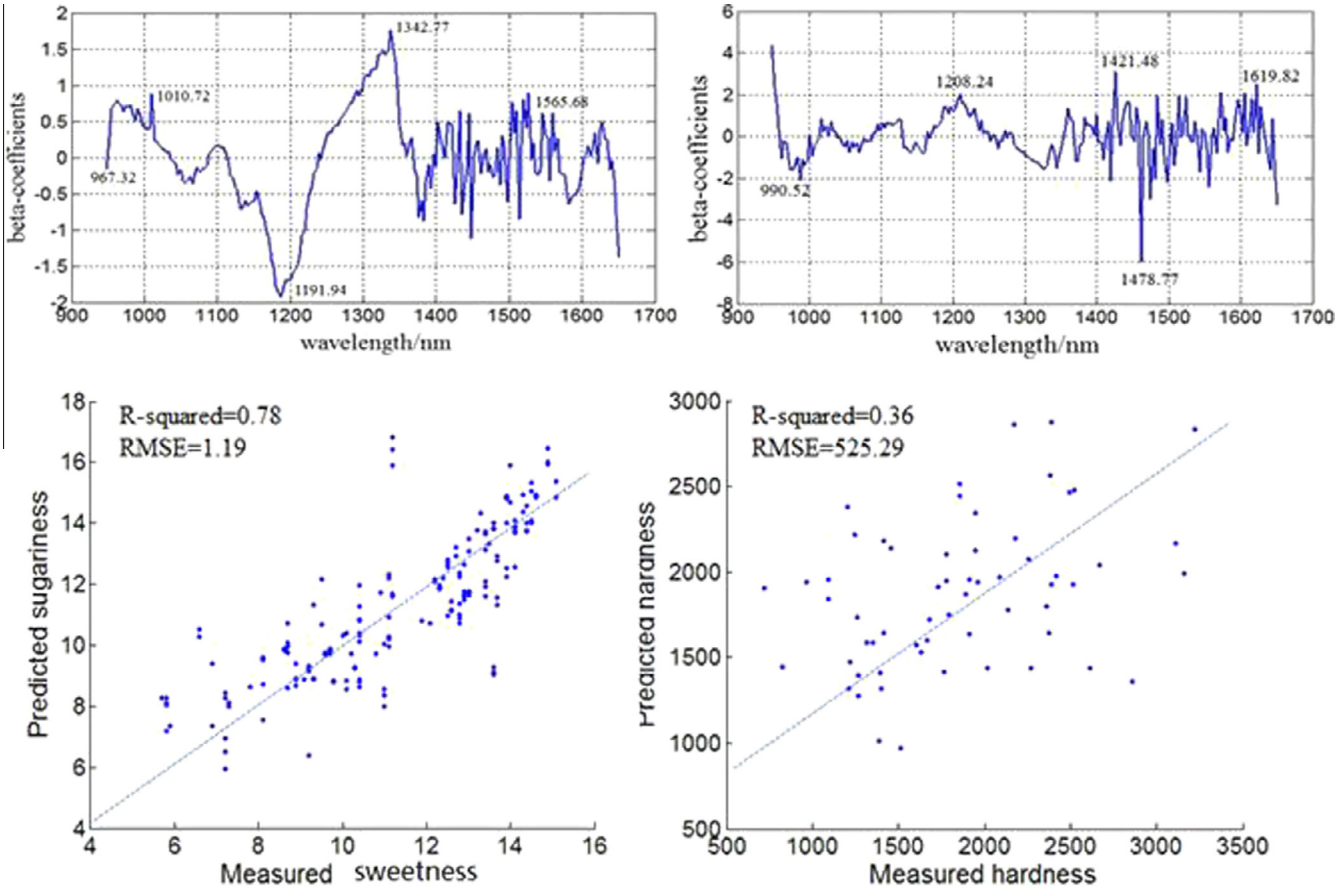

Fig. 5. Selection of important wavelengths for (Up-Left) sweetness and (Up-Right) hardness. Then the predictive result of sweetness and hardness are improved by using the sensitive wavelengths via the PLSR model for (Down-Left) sweetness and (Down-Right) hardness.

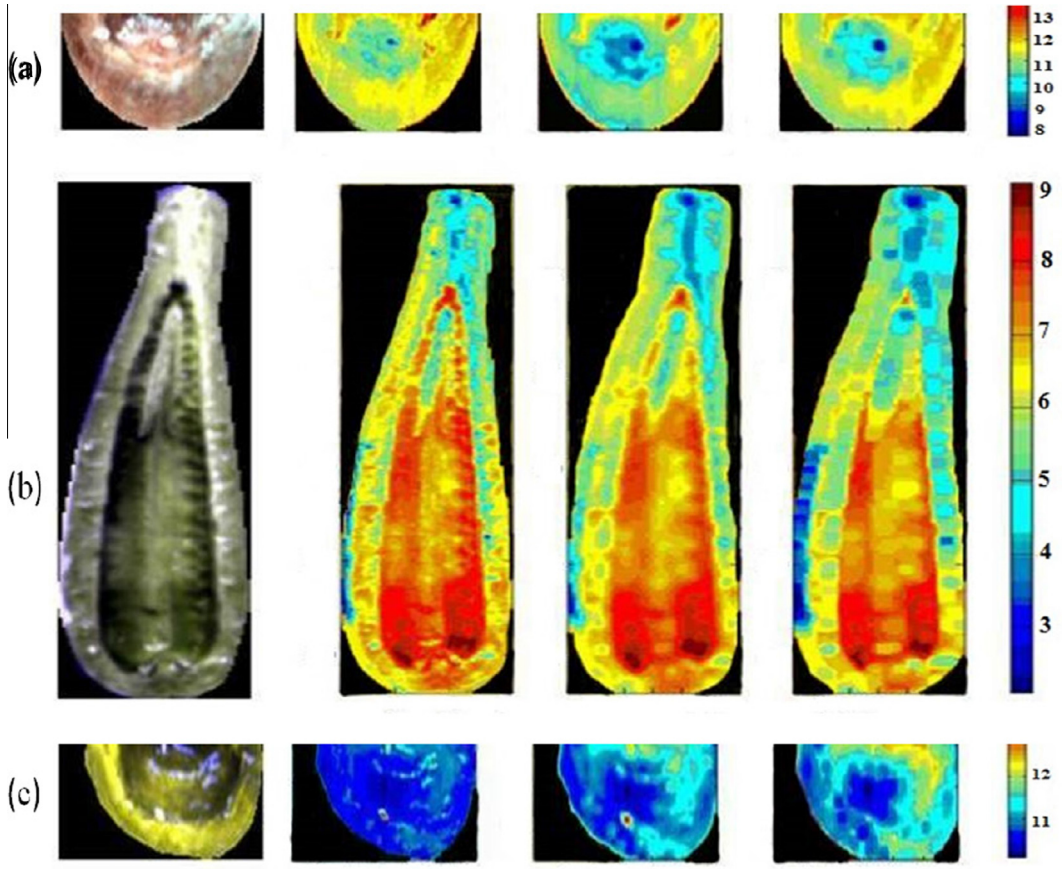

Fig. 6. Distribution maps of sweetness in the melon samples. (a) Sweetness distribution maps of the Elizabeth variety. The first column is the image of the sample that was constructed by concatenating three spectral sub-images at 950 nm, 1200 nm and 1300 nm. The second columns are the sweetness distribution maps. The third column is the result after the mean filter was applied. The fourth column is the result after the median filter was applied. (b) Sweetness distribution maps of the Claw Brittle variety. (c) Sweetness distribution maps of the Emerald variety. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

values of 1.1871 and 525.2932 for sweetness and hardness, respectively. The R-squared values for the Claw Brittle variety were 0.6322 and 0.2999 with RMSEs of 1.1638 and 370.0013 for sweetness and hardness, respectively. The R-squared values of the Emerald variety were 0.5720 and 0.3468 with RMSEs of 1.1729 and 391.0674 for sweetness and hardness, respectively. These results have better accuracy compared with those of the full spectral space. Table 3 shows the primary statistical results of the PLSR models that were developed by using all wavelengths and sensitive wavelengths

3.4. Visual representation of melons

From the above-mentioned quantitative analysis, we determined that the PLSR model has the best performance. Thus, we used the same model on all the tested hyperspectral images to predict all the pixel parameters. The important wavelengths were selected, as mentioned in Section 3.3. The reduced feature information was then input into the best PLSR model to predict sweetness. The PLSR model calculated all the pixels in the reduced spectral space to obtain the predicted parameter values. The resulting image is referred to as the "distribution map" of this attribute. Different colours represent different parameter values. We then observed that this attribute varies from sample to sample or from spot to spot within the same sample. Pixels with similar spectral features produced the same predicted value for this attribute, which was then visualized in the resulting distribution map with a similar colour.

Fig. 6 shows distribution maps of sweetness on the tested samples. In Fig. 6(a), the first column represents the tested samples. The second column represents the sweetness value distribution maps. The third column represents the mean filter-treated

Table 3
PLSR model for predicting sweetness and hardness using the important spectral range.

PLSR model	Full wavelengths		Important wavelengths	
	R-squared	RMSE	R-squared	RMSE
Elizabeth sweetness	0.7952	1.1698	0.7755	1.1871
Elizabeth hardness	0.3671	503.2680	0.3555	525.2932
Claw Brittle sweetness	0.6568	0.6568	0.6322	1.1638
Claw Brittle hardness	0.3417	0.3417	0.2999	370.0013
Emerald sweetness	0.5723	1.1771	0.5720	1.1729
Emerald hardness	0.3412	399.6786	0.3468	391.0674

distribution maps. The last column was obtained after the median filter process. This process was the same for Fig. 6(b) and (c). In the sweetness distribution maps of the Claw Brittle variety shown in Fig. 6(b), the sweetness close to the fleshy part is higher than it is for the part that is near the peel, as expected. The red colour represents higher sweetness, and most of the colour was distributed within the innermost part of the flesh. However, little sweetness is represented by the blue colour, which was close to the peel. This finding is consistent with our everyday experience.

4. Conclusions

In this paper, an NIR hyperspectral imaging system was used to predict melon quality attributes. The results showed that, out of the three prediction models, the PLSR model predicted melon sweetness and hardness by using the spectral information. In addition to predicting the parameters at the sampling point, this non-destructive technique produces distribution maps of the parameters on the tested samples. Furthermore, we determined that it was possible to enhance the results. Thus, in future work, external validations, an analysis model and a larger number of samples will be studied.

Conflict of interest statement

None.

Acknowledgments

The authors wish to acknowledge the support they received from the National Natural Science Foundation, China, under grants 61003201 and 61202165, and the joint project funded by the Royal Society of Edinburgh and NSFC 61211130125.

References

Antonucci, F., Pallottino, F., Paglia, G., Palma, A., D'Aquino, S., & Menesatti, P. (2011). Nondestructive estimation of Mandarin maturity status through portable VISNIR spectrophotometer. *Food and Bioprocess Technology, 4*, 809–813.

Bock, L. E., & Connelly, R. K. (2008). Innovative uses of near-infrared spectroscopy in food processing. *Journal of Food Science, 73*, R91–R98.

Camps-Valls, G., Gomez-Chova, L., & Munoz-Mari, J. (2006). Composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters, 3*, 93–97.

Cariou, C., Chehdi, K., & Le-Moan, S. (2011). BandClust: An unsupervised band reduction method for hyperspectral remote sensing. *IEEE Geoscience and Remote Sensing Letters, 8*, 565–569.

Cen, H. Y., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science and Technology, 18*, 72–83.

Chagnot, C., Venien, A., Jamme, F., Refregiers, M., Desvaux, M., & Astruc, T. (2015). Hyperspectral deep ultraviolet autofluorescence of muscle fibers is affected by postmortem changes. *Journal of Agricultural and Food Chemistry, 69*, 4782–4789.

Cheng, J., & Sun, D. (2015). Data fusion and hyperspectral imaging in tandem with least squares support vector machine for prediction of sensory quality index scores of fish fillet. *LWT – Food Science and Technology, 63*, 892–898.

Desmond, E. M., Kenny, T. A., Ward, P., & Sun, D. W. (2000). Effect of rapid and conventional cooling methods on the quality of cooked ham joints. *Meat Science, 56*, 271–277.

ElMasry, G., Sun, D., & Allen, P. (2012). Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *Journal of Food Engineering, 110*, 127–140.

ElMasry, G., Wang, N., ElSayed, A., & Ngadi, M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Postharvest Biology and Technology, 81*, 98–107.

Foody, G. M., & Mathur, A. (2006). The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment, 103*, 179–189.

Garini, Y., Young, I. T., & McNamara, G. (2006). Spectral imaging, principles and applications. *Cytometry Part A, 69A*, 735–747.

Gowen, A. A., O'Donnell, C. P., & Cullen, P. J. (2007). Hyperspectral imaging – An emerging process analytical tool for food quality and safety control. *Trends in Food Science and Technology, 18*, 590–598.

He, H., & Sun, D. (2015). Hyperspectral imaging technology for rapid detection of various microbial contaminants in agricultural and food products. *Trends in Food Science & Technology, 46*, 99–109.

Heigl, N., Petter, C. H., Lieb, M., Bonn, G. K., & Huck, C. W. (2009). Near infrared reflection spectroscopy and partial least squares regression for determining the total carbon coverage of silica packings for liquid chromatography. *Vibrational Spectroscopy, 49*, 155–161.

Herrero, A. N. (2008). Raman spectroscopy a promising technique for quality assessment of meat and fish: A review. *Food Chemistry, 107*, 1642–1651.

Hou, Y., Song, I., Min, H. K., & Hoon Park, C. (2012). Complexity reduced scheme for feature extraction with linear discriminant analysis. *IEEE Transactions on Neural Networks and learning Systems, 23*, 1003–1009.

Indahl, U., & Naes, T. (2004). A variable selection strategy for supervised classification with continuous spectroscopic data. *Journal of Chemometrics, 18*, 53–61.

Jiang, J., Qiao, X., & He, R. (2016). Use of near-infrared hyperspectral images to identify moldy peanuts. *Journal of Food Engineering, 169*, 284–290.

Keskin, M., Dodd, R. B., Han, Y. J., & Khalilian, A. (2004). Assessing nitrogen content of golf course turfgrass clippings using spectral reflectance. *Applied Engineering in Agriculture, 20*, 851–860.

Kuo, B. C., & Landgrebe, D. A. (2004). Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 1096–1105.

Liao, W., Pižurica, A., Scheunders, P., Philips, W., & Pi, Y. (2013). Semisupervised local discriminant analysis for feature extraction in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing, 51*, 184–198.

Lorente, D., Aleixos, N., Gmez-Sanchis, J., Cubero, S., & Blasco, J. (2013). Selection of optimal wavelength features for decay detection in citrus fruit using the ROC curve and neural networks. *Food and Bioprocess Technology, 6*, 530–541.

Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., Garca Navarrete, O. L., & Blasco, J. (2012). Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology, 5*, 1121–1142.

Ma, J., Sun, D., & Pu, H. (2016). Spectral absorption index in hyperspectral image analysis for predicting moisture contents in pork longissimus dorsi muscles. *Food Chemistry, 197*, 848–854.

McDonald, K., Sun, D. W., & Kenny, T. (2001). The effect of injection level on the quality of a rapid vacuum cooled cooked beef product. *Journal of Food Engineering, 47*, 139–147.

Menon Soumya, V., & Rao, T. V. (2014). Health-promoting components and related enzyme activities of muskmelon fruit during its development and ripening. *Journal of Food Biochemistry, 38*, 415–423.

Mery, D., Pedreschi, F., & Soto, A. (2015). Automated design of a computer vision system for visual food quality evaluation. *Food and Bioprocess Technology, 6*, 2093–2108.

Nogalesbueno, J., Ayala, F., & Heredia, F. (2015). A simplified method for the screening of technological maturity of red grape and total phenolic compounds of red grape skin: Application of the characteristic vector method to near-infrared spectra. *Journal of Agricultural and Food Chemistry, 63*, 4284–4290.

Osborne, S., Jordan, R., & Kunnemeyer, R. (1997). Methods of wavelength selection for partial least squares. *Analyst, 122*, 1531–1537.

Peng, Y., & Lu, R. (2006). Improving apple fruit firmness predictions by effective correction of multispectral scattering images. *Postharvest Biology and Technology, 41*, 266–274.

Prevolnik, M., Candek-Potokar, M., & Skorjanc, D. (2010). Predicting pork water-holding capacity with NIR spectroscopy in relation to different reference methods. *Journal of Food Engineering, 98*, 347–352.

Pu, Y., Feng, Y., & Sun, D. (2015b). Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: A review. *Comprehensive Reviews in Food Science and Food Safety, 14*, 176–188.

Pu, H. B., Sun, D. W., & Ma, J. (2015a). Classification of fresh and frozen-thawed pork muscles using visible and near infrared hyperspectral imaging and textural analysis. *Meat Science, 99*, 81–88.

Qin, J., Burks, T. F., Zhao, X., Niphadkar, N., & Ritenour, M. A. (2012). Development of a two-band spectral imaging system for real-time citrus canker detection. *Journal of Food Engineering, 108*, 87–93.

Sinija, V., & Mishra, H. (2011). FTNIR spectroscopic method for determination of moisture content in green tea granules. *Food and Bioprocess Technology, 4*, 136–141.

Theron, L., Venien, A., Jamme, F., Fernandez, X., Peyrin, F., Molette, C., et al. (2015). Protein matrix involved in the lipid retention of foie gras during cooking: A multimodal hyperspectral imaging study. *Journal of Agricultural and Food Chemistry, 62*, 5954–5962.

Wang, W., Li, C., Tollner, E. W., Gitaitis, R. D., & Rains, G. C. (2012). Shortwave infrared hyperspectral imaging for detecting sour skin (*Burkholderia cepacia*)-infected onions. *Journal of Food Engineering, 109*, 38–48.

Zhang, B. H., Huang, W. Q., & Li, J. B. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Research International, 62*, 326–343.