

RAPID DISCRIMINATION OF HIGH-QUALITY WATERMELON SEEDS BY MULTISPECTRAL IMAGING COMBINED WITH CHEMOMETRIC METHODS

Wei Liu^{1,2}, Xue Xu³, Changhong Liu^{1*}, Lei Zheng^{1*}

¹ School of Food Science and Engineering, Hefei University of Technology, Hefei 230009, China; e-mail: liuchanghong1982@163.com

² Intelligent Control and Compute Vision Lab, Hefei University, Hefei 230601, China

³ Rice Research Institute, Anhui Academy of Agricultural Sciences, Hefei 230031, China

This study focuses on the feasibility of nondestructive discrimination of high-quality watermelon seeds with a multispectral imaging system combined with chemometrics. Principal component analysis (PCA), least squares-support vector machines (LS-SVM), back propagation neural network (BPNN), and random forest (RF) were applied to determine the seed quality. The results demonstrate that both the spectral and the morphological features are essential for discrimination of the quality of watermelon seeds. Clear differences between high-quality watermelon seeds and other watermelon seeds including dead seeds and low-vigor seeds were visualized, and an excellent classification (with accuracies of 92% in the LS-SVM model for Julong and 91% in the RF model for Xiali, respectively) was achieved. These results indicate that multispectral imaging could be used for rapid and efficient non-destructive quality control of watermelon seeds.

Keywords: watermelon seeds, multispectral imaging, nondestructive.

БЫСТРАЯ ИДЕНТИФИКАЦИЯ КАЧЕСТВА СЕМЯН АРБУЗА С ПОМОЩЬЮ МУЛЬТИСПЕКТРАЛЬНОГО ПРЕДСТАВЛЕНИЯ В СОЧЕТАНИИ С ХЕМОМЕТРИЧЕСКИМИ МЕТОДАМИ

W. Liu^{1,2}, X. Xu³, Ch. Liu^{1*}, L. Zheng^{1*}

УДК 519.281.2;635.615/618

¹ Школа пищевых технологий, Университет технологии Хэфэя, Хэфэй 230009, Китай; e-mail: liuchanghong1982@163.com

² Университет Хэфэя, Хэфэй 230601, Китай

³ Исследовательский институт риса, Аньхойская академия сельскохозяйственных наук, Хэфэй 230031, Китай

(Поступила 26 декабря 2017)

Рассмотрена возможность осуществления неразрушающего контроля качества семян арбуза, основанного на использовании их мультиспектральной визуализации в сочетании с хемометрикой. Для определения качества семян предложено использовать анализ основных компонент (PCA), метод наименьших квадратов – опорных векторов (LS-SVM), алгоритм нейронной сети с обратным распространением ошибки (BPNN) и модель случайного леса (RF). Показано, что как спектральные, так и морфологические данные являются ключевыми факторами для определения качества семян арбуза. Различие между высоко- и низкокачественными (мертвыми, со слабой всхожестью) семенами арбуза может быть визуализировано и достаточно точно идентифицировано (до 92% с помощью модели LS-SVM для сорта Julong и 91% с помощью метода RF для сорта Xiali).

Ключевые слова: семена арбуза, мультиспектральное представление, неразрушающий контроль.

Introduction. Watermelon (*Citrullus lanatus*) is a crop accounting for 7% of the worldwide area related to vegetable production [1], highly important due to its nutritional value and taste [2, 3]. The quality of watermelon seeds, especially their purity and viability, affects all the production stages, such as sowing, growing, and harvesting. Low-quality seeds can include seeds of other varieties affected by cross-pollination and nonviable seeds influenced by poor storage conditions [4]. Thus, the discrimination of high-quality watermelon seeds has become critical to the development of the watermelon seed market.

A few methods have been developed to discriminate high-quality seeds, namely grow-out-trial (GOT) to assess the seed purity and the viability in a crop variety [5, 6]; isozyme electrophoresis technologies for seed genetic purity testing [7]; and tetrazolium or electrical conductivity tests to evaluate the seed viability [8]. However, these methods are destructive, time-consuming, costly, and often influenced by the environment, which may affect the accuracy of discrimination [9]. Recently, some rapid and nondestructive methods have been proposed to resolve this problem, such as near infrared reflectance spectroscopy and Raman spectroscopy in the estimation of components and seed viability [10, 11], LED-induced hyperspectral reflectance imaging to determine the germination quality of cucumber seeds [12], hyperspectral imaging to measure the seed viability of corn [13], and the multispectral fluorescence imaging technique to discriminate the cucumber seed viability [14].

Multispectral imaging is a developing nondestructive technology that integrates the benefits of conventional imaging and spectroscopy to simultaneously achieve both spatial and spectral information from the target object. Analysis from multispectral imaging is simple, rapid, nondestructive, and does not require sample pre-treatment, so this technique is well suited for on-line process monitoring and quality control [15, 16]. Recently, multispectral imaging has been applied to predict the quality and maturity of fruit and vegetables [17–19], detect the fungi infection in corn kernels [20], detect food adulteration [21, 22], and discriminate transgenic rice seeds [23]. However, there are no published data on multispectral imaging for discrimination of high-quality watermelon seeds. Therefore, the aim of this research is to assess the potential of the multispectral imaging technique for determining high-quality watermelon seeds used in combination with chemometrics methods, including least squares-support vector machines (LS-SVM), back propagation neural network (BPNN), and random forest (RF).

Calculation. The two varieties of hybrid watermelon seeds (Xiali and Julong, 500 samples for each variety) used in the experiment were provided by the Anhui Jianghuai Horticulture Seeds Co., Ltd, Hefei, China. All the samples were labeled and stored in sealed plastic bags at 4°C until use.

Image acquisition and analysis. A VideometerLab (Videometer A/S, Hørsholm, Denmark) was used to collect the multispectral image, which included 19 different wavelengths (405, 435, 450, 470, 505, 525, 570, 590, 630, 645, 660, 700, 780, 850, 870, 890, 910, 940, and 970 nm). The principal setup of the system was composed of a point-grey scorpion camera, light emitting diodes (LEDs), and an integrating sphere with a matte white coating used to guarantee that the light is uniform, diffuse, and evenly scattered. At the rim of the sphere, LEDs were positioned side by side so as to distribute the LEDs at the specific wavelength uniformly around the entire rim. The LEDs were strobing successively, resulting in an image of 1280×960 for each LED.

The multispectral system was firstly calibrated with both a diffuse white and a dark target to ensure pixel correspondence for all spectral bands. The object was placed inside the integrating sphere, and then a multispectral image consisting of 19 separate images was recorded. The spectra in the range 405–970 nm, including the visible and the lower wavelengths of the NIR region, were collected from each sample image. Each multispectral image was processed with VideometerLab software (version 2.12.23). The image background was removed by canonical discriminant analysis (CDA) and segmented using a simple threshold. The important morphological features of each watermelon seed, including area (mm²), length (mm), width (mm), and Hunter L^* , a^* , b^* values, were all extracted from the image analysis and processing.

After recording the multispectral images, the samples were transported to the breeding base for seeds in Hainan, South China, to determine their quality by the GOT method.

Principal component analysis (PCA). As a common unsupervised recognition method, PCA could reduce the original reflectance spectra to a smaller number of variables, called principal components (PC). In this study the dimensions of the data matrix from the seed samples were reduced and the main information was extracted. The final results of PCA, consisting of score plots, could be used to visualize the contrast between the watermelon seeds and provide essential information about differentiating the samples.

LS-SVM was used as a learning algorithm for classification and regression tasks [24]. A radial basis function (RBF) with the Gaussian function was selected as the kernel function to reduce the computational

complexity of the training procedure and give good performance under the general smoothness assumptions. Two crucial parameters (γ , σ^2) were needed for LS-SVM, and higher performance of the model can be obtained by adjusting the value of the parameters. The details of the LS-SVM algorithm can be found in the previously reported researches. Here, to find the optimum parameters of the model with the best discrimination results, leave-one-out cross-validation was performed during the calibration step, and the selection ranges of γ and σ^2 were defined from 2^{-10} to 2^{10} [25].

BPNN can solve complex problems more accurately than the linear techniques, so it has been widely used for pattern recognition in many fields. A three-layer structure BPNN (an input layer, a hidden layer, and an output layer) was selected. Leave-one-out cross-validation was performed during the calibration step. Several network architectures were tested by varying the number of neurons in the hidden layer with different initial weights. The optimal parameters (hidden nodes, the goal error, and iteration times) were determined by the least prediction error.

RF, as a combination of tree predictors, is one of the most successful classifiers based on the ensemble learning algorithm, which shows high resistance to noisy variables and can handle a large number of variables [26, 27]. RF is an ensemble classification algorithm that consists of many classification and regression trees. The leaves of classification trees represent class labels, and the branches represent conjunctions of the features that lead to class labels. Here, ensemble B trees are presented as $\{T_1(x), \dots, T_B(x)\}$, where $x \in \{x_1, \dots, x_m\}$ is the m -dimensional vector of the variables of a classified object, which is the m -dimensional vector of the features obtained from the multispectral image of the samples. The outputs of ensemble B are produced as $\{\hat{y}_1 = T_1(x), \dots, \hat{y}_B = T_B(x)\}$, where \hat{y}_b , $b \in \{1, 2, \dots, B\}$ is the prediction for a classified object by the b th tree. The outputs of all trees are aggregated to produce one final prediction \hat{y} .

Results and discussion. *Analysis of reflectance spectra.* From the results of the GOT method, all the seed samples were divided into four classes, which were pure seeds, seeds of other varieties, dead seeds, and low-vigor seeds. For 500 Julong seeds, 291 were pure seeds, 109 were seeds of other varieties, 40 were dead seeds, and 60 were low-vigor seeds. For 500 Xiali seeds, 278 were pure seeds, 111 were seeds of other varieties, 53 were dead seeds, and 58 were low-vigor seeds. The spectral images of the four classes of watermelon seeds are shown in Fig. 1.

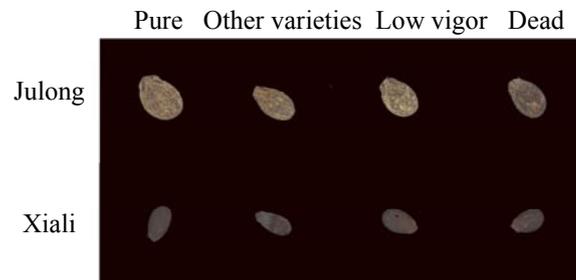


Fig. 1. Images of four classes of watermelon seeds.

The average reflectance spectra of two varieties of watermelon seeds (Julong and Xiali) in the wavelength range 405–970 nm are shown in Fig. 2. The spectral reflectance curves of the samples were smooth, and the general trend of all spectra was similar between the two varieties of watermelon seeds. There was some subtle difference between pure seeds and the other three classes of seeds, especially from 600 to 900 nm, which may be due to the differences of the color or chemical components of the seed samples. But as for the low-vigor seeds, it was not easy to distinguish them from the other classes, which may be due to the uncertainty in seed variety.

Here, all the pure and high-vigor seeds were regarded as high-quality seeds, and to simplify the problem of high-quality seed discrimination, all the seed samples were divided into two classes: high-quality seeds and other seeds (including seeds of other varieties, dead seeds, and low-vigor seeds). A total of 200 high-quality seeds and 100 other seeds was randomly selected from the samples and then divided into the calibration set (including 140 high-quality seeds and 60 other seeds) and the prediction set (including 60 high-quality seeds and 40 other seeds). The average morphological values of watermelon seeds, including area (mm^2), length (mm)/width (mm), roundness, and Hunter L^* , a^* , b^* , are shown in Table 1. From Table 1, we can see that there is clear difference between the high-quality seeds and other seeds in such morphological parameters as roundness, CIE a^* , and area.

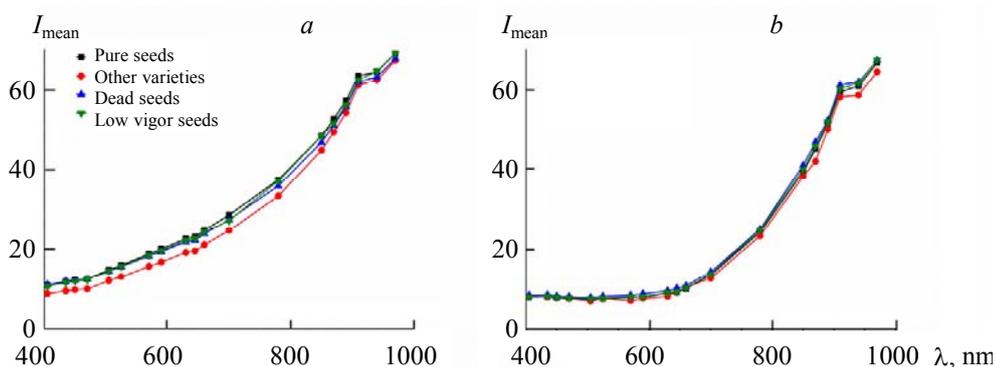


Fig. 2. Average spectra from the multispectral images of high-quality and other watermelon seeds. (a) Julong variety; (b) Xiali variety.

TABLE 1. Morphological Features of Watermelon Seeds of Two Varieties

Variety	Class	Area, mm ²	Length/Width, mm	Roundness, mm	CIE <i>L</i> *	CIE <i>a</i> *	CIE <i>b</i> *
Julong	High-quality	60.573	11.189	7.093	2.083	49.089	5.930
	Other	69.302	11.676	7.864	2.042	45.113	6.072
Xiali	High-quality	61.267	11.245	7.125	2.151	48.578	6.164
	Other	65.426	11.538	7.493	2.096	47.724	6.257

PCA was performed initially to examine the qualitative difference of the high-quality seeds and other seeds in the PC space with the combined spectral and morphological features data, and the results can be found in Fig. 3. The three-dimensional principal component (PC) score plot was obtained with the first three score vectors (PC1, PC2, PC3) derived from the combined spectral and morphological features data. From the PCA plot, the first three PCs, accounting for the most variation, were 99.64% and 98.85% for Julong and Xiali, respectively. However, there was no apparent differentiation between the high-quality seeds and other seeds in both watermelon varieties with the combined spectral and morphological features data. Thus, it is difficult to discriminate the high-quality watermelon seeds with simple linear methods, so other nonlinear methods, including LS-SVM, BPNN, and RF, were utilized for improved separation.

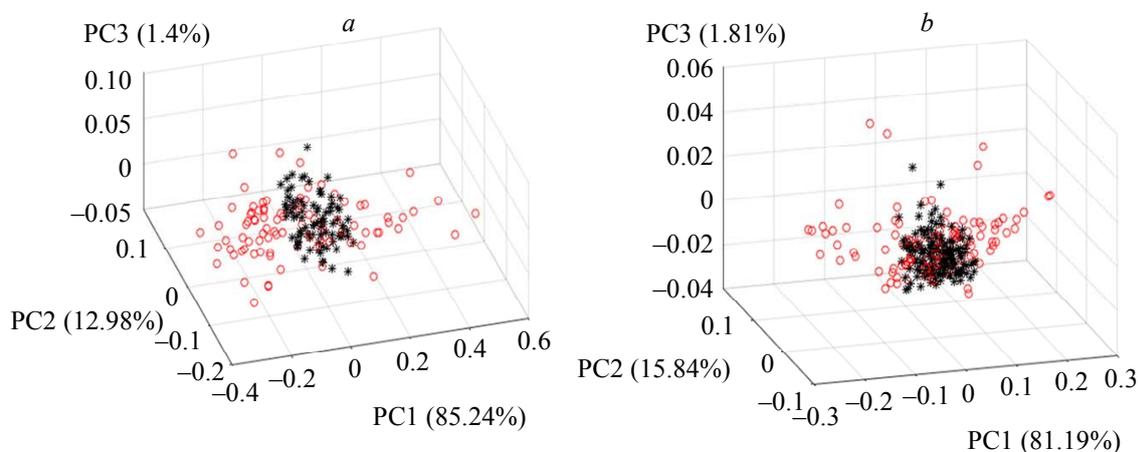


Fig. 3. Three-dimensional score plot of the first three principal components for the high-quality (*) and other (o) watermelon seeds with the combined spectral and morphological features data. (a) Julong variety; (b) Xiali variety.

Discrimination of high-quality watermelon seeds. Discrimination models for the two varieties of watermelon seeds based on the spectral features data or the combined spectral and morphological features data were established using LS-SVM, BPNN, and RF, respectively. The results can be found in Table 2. For the Julong variety samples, all the models with the combined spectral and morphological features data were better than the models with the spectral features only. At the model development stage using LS-SVM with the RBF kernel, the key parameters (γ, σ^2) that determined the boundary complexity and the prediction performance were found to be (64, 0.00515433) in the combined spectral and morphological features data using cross validation, which is shown in Fig. 4a. The accuracies of the developed model were 94% in the calibration set and 92% in the prediction set, respectively. In the process of the model development with BPNN using the combined spectral and morphological features data, the parameters including hidden nodes, the goal error, and iteration times, were determined to be 20, 1×10^{-8} , and 800, respectively. The results showed that the accuracies of the model in the calibration and prediction sets were 88.5% and 84%, respectively. In the process of the model development with RF using the combined spectral and morphological features data, the number of classification trees desired (n_{tree}) was defined as 25, and the number of variables (m_{try}) used in each tree to make the tree grow was also 25. The accuracies in the calibration and prediction sets were 98% and 87%, respectively.

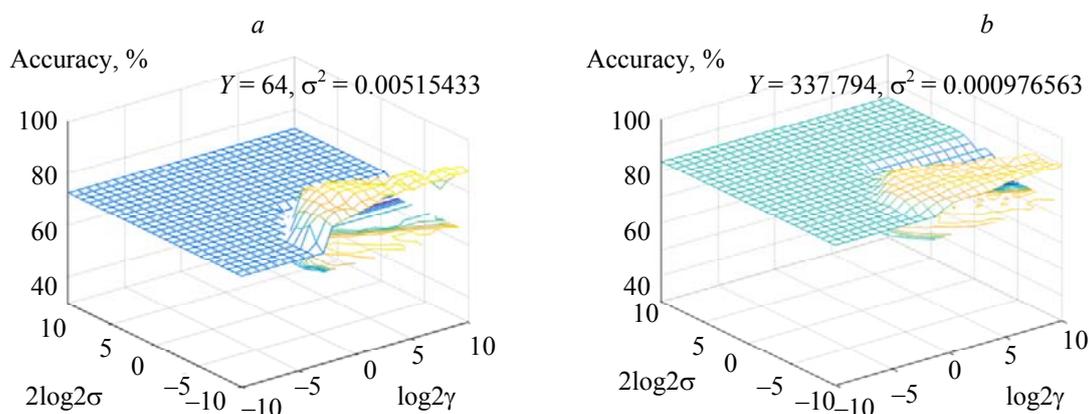


Fig. 4. Plot of tuning of γ and σ^2 for LS-SVM model of Julong variety samples (a) and Xiali variety samples (b).

TABLE 2. Comparison of Discrimination Performance Obtained with LS-SVM, BPNN, and RF methods, with the Spectral Data and the Combined Spectral and Morphological Features Data

Sample	Chemometric methods	Features	Calibration set		Prediction set	
			MS	Accuracy, %	MS	Accuracy, %
Julong	LS-SVM	Spectral	16	92	13	87
		Spectral + morphology	12	94	8	92
	BPNN	Spectral	31	84.5	24	76
		Spectral + morphology	23	88.5	16	84
RF	Spectral	8	96	21	79	
	Spectral + morphology	4	98	13	87	
Xiali	LS-SVM	Spectral	15	92.5	21	79
		Spectral + morphology	17	91.5	17	83
	BPNN	Spectral	17	91.5	28	72
		Spectral + morphology	12	94	25	75
RF	Spectral	8	96	19	81	
	Spectral + morphology	9	95.5	9	91	

Note. MS is misclassified samples.

Similar to the Julong variety, the models with the combined spectral and morphological features data were also better than the models only with the spectral features for Xiali variety samples. The optimal results were obtained by the LS-SVM model with the parameters (γ , σ^2) at (337.794, 0.000976563), which can be seen in Fig. 4b. The accuracies of the developed model were 91.5% and 83% in the calibration and prediction sets, respectively. In the BPNN model, the optimal parameters were the same as the Julong samples. Compared to the LS-SVM method, the results were worse in the prediction set with an accuracy of only 75%. In the model development with RF, the parameters were also the same as with the Julong samples. The accuracies in the calibration and prediction sets were 95.5% and 91%.

From Table 2, the discrimination performances with different chemometric methods were also compared, and the accuracy results obtained from the calibration and prediction sets were summarized accordingly. Detailed comparison based on the accuracies showed that LS-SVM and RF were the best chemometric methods for determining high-quality watermelon seeds with the combined spectral and morphological features data for the Julong and Xiali varieties, respectively. The results in the prediction set can be found in Fig. 5. In the prediction set, the misclassified numbers were only 8 for Julong and 9 for Xiali, and the respective accuracies were 92% and 91% with the LS-SVM and RF methods, respectively.

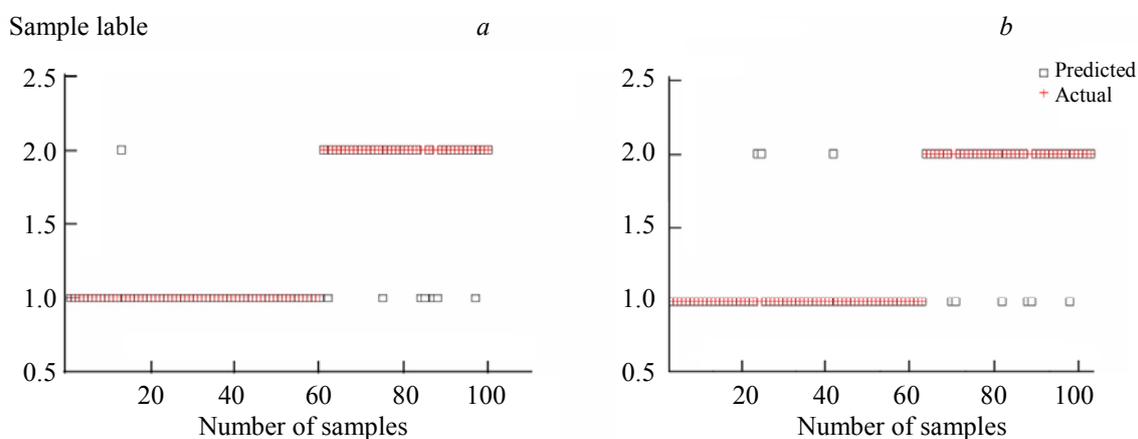


Fig. 5. The results of quality discrimination obtained using different chemometrics in the prediction set. (a) LS-SVM for Julong variety; (b) RF for Xiali variety.

Conclusion. The spectral reflectance and morphological features obtained from a multispectral imaging system have been considered as important features for determining high-quality watermelon seeds. The LS-SVM model using the combined spectral and morphological features data is shown to have the best prediction ability, with an accuracy of 92% in the Julong variety. Meanwhile, in the Xiali variety, the best model is obtained using the RF method and the combined spectral and morphological features data, with an accuracy of 91%. The encouraging results demonstrate that the use of the multispectral imaging technique combined with chemometric methods has the potential to be widely used for rapid and on-site determination of high-quality watermelon seeds.

Acknowledgment. This work was supported by the National Key Research and Development Plan of China (2016YFD0401104), the Key Science & Technology Specific Projects of Anhui Province (15czz03117), the Fundamental Research Funds for the Central Universities (JZ2016HG TB0712), the Talent Research Fund of Hefei University (15RC09), and the Natural Science Research Key Program of Anhui Advanced Education (KJ2015A258).

REFERENCES

1. A. Guo, J. Zhang, H. Sun, J. Salse, W. J. Lucas, *Nat. Genet.*, **45**, 51–58 (2012).
2. Perkins-Veazie, P. Collins, J. K. Davis, A. R. Roberts, *J. Agric. Food Chem.*, **54**, 2593–2597 (2006).
3. J. K. Collins, G. Y. Wu, P. Perkinsveazie, K. Spears, P. L. Claypool, R. A. Baker, *Nutr.*, **23**, 261–266 (2007).
4. D. S. Egel, C. Gunter, R. Martyn, *Hortscience*, **43**, 1410–1414 (2008).

5. K. K. Moorthy, P. Babu, M. Sreedhar, V. S. A. K. Sama, P. N. Kumar, S. M. Balachandran, R. M. Sundaram, *Seed Sci. Technol.*, **39**, 282–292 (2011).
6. S. Ye, Y. Wang, D. Huang, J. Li, Y. Gong, L. Xu, L. Liu, *Sci. Hort-Amsterdam*, **155**, 92–96 (2013).
7. M. D. Sánchez-Yélamó, *Seed Sci. Technol.*, **34**, 297–306 (2006).
8. V. N. Soares, S. G. Elias, G. I. Gadotti, A. E. Garay, F. A. Villela, *Crop Sci.*, **56**, 707–715 (2016).
9. E. Noli, S. Conti, M. Maccaferri, M. C. Sanguineti, *Seed Sci. Technol.*, **27**, 1–10 (1999).
10. N. Shetty, T. G. Min, R. Gislum, M. Olesen, B. Boelt, *J. Near Infrared Spectrosc.*, **19**, 451 (2012).
11. S. S. Shin, E. J. Shin, D. P. Sang, K. S. Lee, *Can. J. Civil. Eng.*, **16**, 538–546 (2012).
12. C. Mo, J. Lim, K. Lee, S. Kang, M. S. Kim, G. Kim, *J. Biosyst. Eng.*, **38**, 318–326 (2013).
13. A. Ambrose, S. Lohumi, W. H. Lee, B. K. Cho, *Sensor Actuator B: Chem.*, **224**, 500–506 (2016).
14. C. Mo, M. S. Kim, J. Lim, K. Lee, G. Kim, B. K. Cho, *Trans. ASABE*, **58**, 959–968 (2015).
15. Y. Z. Feng, D. W. Sun, *Crc. Crit. Rev. Food Sci. Nutr.*, **52**, 1039–1058 (2012).
16. J. Qin, K. Chao, M. S. Kim, R. Lu, T. F. Burks, *J. Food Eng.*, **118**, 157–171 (2013).
17. F. A. E. Hahn, *Biosyst. Eng.*, **81**, 147–155 (2002).
18. M. M. Løkke, H. F. Seefeldt, T. Skov, M. Edelenbos, *Posthavrvest Biol. Technol.*, **75**, 86–95 (2013).
19. C. Liu, W. Liu, W. Chen, J. Yang, L. Zheng, *Food Chem.*, **173**, 482–488 (2015).
20. T. C. Pearson, D. T. Wicklow, *Trans. ASABE*, **49**, 1235–1245 (2006).
21. A. I. Ropodi, D. E. Pavlidis, F. Mohareb, E. Z. Panagou, G. J. Nychas, *Food Res. Int.*, **67**, 12–18 (2015).
22. A. I. Ropodi, E. Z. Panagou, G. J. E. Nychas, *Food Control*, **73**, 57–63 (2016).
23. C. Liu, W. Liu, X. Lu, W. Chen, J. Yang, L. Zheng, *Food Chem.*, **153**, 87–93 (2014).
24. C. Cortes, V. Vapnik, *Mach. Learn.*, **20**, 273–297 (1995).
25. O. Devos, C. Ruckebusch, A. Durand, L. Duponchel, J. P. Huvenne, *Chemometer Intell. Lab. Syst.*, **96**, 27–33 (2009).
26. U. Bradter, T. J. Thom, *J. Appl. Ecol.*, **48**, 1057–1065 (2011).
27. M. Liu, M. Wang, J. Wang, D. Li, *Sensor Actuator B: Chem.*, **177**, 970–980 (2013).