

Determination of damaged wheat kernels with hyperspectral imaging analysis

Yuanyuan Shao^{1,2}, Chong Gao¹, Guantao Xuan^{1,3*}, Xuemei Gao², Youqing Chen², Zhichao Hu^{2*}

(1. College of Mechanical and Electrical Engineering, Shandong Intelligent Engineering Laboratory of Agricultural Equipment, Shandong Agricultural University, Taian 271018, Shandong, China;

2. Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China;

3. College of Agriculture, Food & Natural Resources, University of Missouri, Columbia, Missouri 65203, USA)

Abstract: Hyperspectral imaging was applied to classify the damaged wheat kernels and healthy kernels. The spectral information was extracted from damaged wheat kernels and healthy kernels samples. The effective wavelengths were obtained from spectral of 865-1711 nm by X-loadings of principal component analysis (PCA) and successive projection algorithm (SPA) method, respectively. Partial least square method (PLS) and least square-support vector machine (LS-SVM) were then used to build classification models on full spectral data and effective wavelengths dataset, respectively. The results showed that the classification accuracy of every LS-SVM model was the best, being 100%. While the accuracy of the PLS model was slightly lower, still over 97%. The confusion matrix showed that several damaged wheat kernels samples were misclassified as healthy samples, while all healthy samples were correctly classified. The overall results indicated that hyperspectral imaging could be used for discriminating the damaged wheat kernels and could provide a reference for detecting other grain kernels grading degrees. Further, this study can provide a research basis for the development of online or portable detectors on grain damaged kernels recognition, which will be beneficial for grain grading or post-harvest quality processing of other grains.

Keywords: hyperspectral image, damaged wheat kernels, determination, PCA, SPA, LS-SVM

DOI: 10.25165/j.ijabe.20201305.4413

Citation: Shao Y Y, Gao C, Xuan G T, Gao X M, Chen Y Q, Hu Z C. Determination of damaged wheat kernels with hyperspectral imaging analysis. *Int J Agric & Biol Eng*, 2020; 13(5): 194–198.

1 Introduction

Wheat is one of the major crops and one of the three major staple foods in China. It is of great importance to classify the quality of wheat for the purpose of ensuring food security, realizing quality prices in wheat market transactions. According to wheat national standards of China^[1], damaged wheat kernel is one important degrading factor in classifying wheat grade. The wheat grade depends on the proportion of damaged wheat kernels, grade 1 and 2 requires the proportion is less than or equal to 6%, while grade 3, 4 requires the proportion is less than 8%, and grade 5 requires less than or equal to 10%^[1]. The damaged wheat kernels normally are caused by insect infestation, mechanical damage, diseased kernels, germinal kernels and moldy kernels during harvest and storage. Such damages seriously affect the quality of food and the safe storage of wheat. Conventionally, damaged

wheat kernels are detected by manual method, which is time-consuming, inefficient, and being subjective, and so on^[2].

As a fast and non-destructive method, the spectroscopy technique has been widely used in crop quality detection. A visible/near-infrared hyperspectral imaging system was used to classify aflatoxin B1 contaminated maize kernels, principal component analysis (PCA) and factorial discriminant analysis (FDA) models showed that the prediction accuracy was over 96%^[3]. Hyperspectral imaging was used to predict protein content in single wheat kernels by building partial least squares (PLS) regression models^[4]. Near-infrared (NIR) hyperspectral imaging system was used to detect five concentration levels of Ochratoxin A (OTA) in contaminated wheat kernels, and PCA was applied to identify the key wavelengths^[5]. Near-Infrared hyperspectral imaging was used to evaluate the diffusion of water into single wheat kernel sections over time with PCA and a supervised method based on the Non-Negative Least Squares (NNLS) algorithm^[6]. Near-infrared hyperspectral imaging was used to detect fungal infection and Ochratoxin A contamination in stored wheat and PCA was applied to obtain significant wavelengths. All the three classifiers of linear, quadratic and Mahalanobis discriminant classifiers differentiated healthy kernels from fungal-infected kernels with a classification accuracy higher 90%^[7]. Besides the application on crop quality detection, spectroscopy technology had been widely used on qualitative analysis of variety^[8,9], adulteration^[10-14], crops of different years^[15], etc.

In terms of identifying damaged wheat kernels, a lot of research has been conducted. The near-infrared hyperspectral imaging was used to detect insect-damaged wheat kernels. In the result, linear discriminant analysis and quadratic discriminant

Received date: 2018-05-20 **Accepted date:** 2020-06-10

Biographies: Yuanyuan Shao, PhD, Associate Professor, research interest: agricultural mechanization, Email: xoyo111@163.com; Chong Gao, Master degree candidate, research interest: agricultural mechanization, Email: 1092601973@qq.com; Xuemei Gao, MS, Assistant Professor, research interest: agricultural machinery design, Email: 591882839@qq.com; Youqing Chen, MS, Associate Professor, research interest: agricultural mechanization, Email: 89081229@qq.com.

***Corresponding author:** Guantao Xuan, PhD, Associate Professor, research interest: agricultural mechanization. Shandong Agricultural University, Taian 271018, Shandong, China. Tel: +86-538-8249831, Email: xuangt@sina.com; Zhichao Hu, PhD, Professor, research interest: agricultural mechanization engineering. Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China. Tel: +86-25-84346246, Email: zchu369@163.com.

analysis classifiers correctly classified 85%-100% healthy and insect-damaged wheat kernels^[16]. The short-wave near-infrared hyperspectral and digital color imaging were used to identify insect-damaged wheat kernels from healthy wheat kernels, and the quadratic discriminant analysis classifier gave the highest accuracy and correctly identified 96.4% healthy and 91.0%-100.0% insect-damaged^[17]. The hyperspectral image was used to detect unsound wheat kernel with a multi-classification support vector machine quickly and accurately, and the total identification rate of the 4 classification models was over 94%^[18].

A color machine vision system was used for the identification of healthy and six types of damaged kernels (broken, grass-green/green-frosted, black-point/smudged, mildewed, heated and bin-/fire-burnt). By using a non-parametric classifier with a selected combined feature model of 24-color and four morphological features, the average classification accuracies were over 90%^[19]. The soft X-ray imaging was used to detect fungal infected wheat from healthy wheat, and several models were built to compare classification accuracies^[20].

Though the above researches mainly focused on the identification of damaged kernels, one kernel was taken as one sample, and different types of damaged wheat kernels were identified. It is exhausting and time-consuming to pick out every damaged wheat kernel of different damaged degrees and collect corresponding spectral information one by one. And the wheat grade only depended on the proportion of damaged wheat kernels (mixed by all types of damaged kernels), not on the single damaged kernel. Thus, it's urgent to study on identifying damaged wheat kernels with piles of wheat kernels and search for one efficient and non-destructive method with hyperspectral imaging technology.

The purposes of this study were to: (1) collect wheat kernels samples and group them into two classifications manually, (2) acquire spectral imaging data of wheat kernels by the hyperspectral imaging system, (3) analyze effective wavelength using X-loadings of PCA and SPA method, (4) establish PLS and LS-SVM models to identify the unhealthy wheat kernels from the healthy ones based on full spectral data and selected effective wavelengths, respectively.

2 Materials and methods

2.1 Experimental materials

The wheat seeds material (Shannong 12) was bought in Seed Company of Shandong Agricultural University in Taian city, Shandong Province, China. The seeds were harvested in Taian city in June 2017. According to the quality of wheat kernels, two categories of samples were classified: damaged and healthy. Samples of three types of damaged kernels including broken, black-point, wheat coat wrinkle were manually picked, as shown in Figure 1. As shown in Figure 1a, there were black points on wheat kernels, which might be caused by fungi and bacteria disease. Figure 1b showed some incomplete wheat kernels, which were broken during harvest or storage. As shown in Figure 1c, there were some wrinkles on the wheat kernels coat, which might be caused by moisture. Figure 1d showed healthy wheat kernels, which were bright and smooth. The damaged and healthy kernels were divided into 360 trays (diameter of 550 mm, depth of 100 mm) respectively. All the trays were full of wheat kernels, and the surface was pushed to flat, as shown in Figure 2. There were 360 samples at each category, and a total of 720 samples were collected in this study.



Figure 1 Examples of damaged and healthy wheat kernels



a. Damaged kernels b. Healthy kernels

Figure 2 Samples of damaged wheat and healthy wheat

2.2 Acquisition of hyperspectral images

2.2.1 Hyperspectral imaging system

In this study, a hyperspectral imaging system (900-1700 nm) with a spectral resolution of 5 nm was used. This system was composed of an imaging spectrograph (ImSpector N17E, Spectral Imaging Ltd., Oulu, Finland), 256 CCD camera (Xeva 992; Xenics Infrared Solutions, Leuven, Belgium) with a camera lens (OPCA05G, Hamamatsu, Japan), computer and electronically controlled mobile platform devices (PSA200-11-X, Zolix Instruments Co., Ltd., Beijing, China), etc. The system was placed in a dark room, four 100W halogen lamps (HSIA-LS-TAIF, Zolix instruments Co., Ltd., Beijing, China) were symmetrically placed at a 45° angle for illumination.

2.2.2 Image acquisition and calibration

Nine trays of samples were evenly placed on the conveyor three by three every time, and the parameters were set as follows: the scan speed was 19 mm/s, the distance between the lens and conveyor was 250 mm, the exposure time of the camera was 11 ms. The NIR hyperspectral images were collected with SpectralSENS (Spectral Imaging Ltd., Finland) and processed using ViewSpecPro6.2.0 and Envi4.6 (Environment for Visualizing Images software, Research Systems Inc., Boulder, CO, USA).

To acquire a hyperspectral image under the same experimental conditions, the dark reference image and the white reference image should be captured for image calibration. The dark reference image was captured by turning off the illumination and covering the lens of the camera with its opaque cap completely. And the white reference image was captured using a standard Teflon white board. Then the calibration was conducted by Equation (1)^[21].

$$I_0 = \frac{I_{raw} - I_{dark}}{I_{white} - I_{dark}} \quad (1)$$

where, I_0 is the calibration image; I_{raw} is the raw image, I_{dark} is the dark reference image; I_{white} is the white reference image.

A region of interest (ROI) with a size of 100×100 pixels in the center of each corrected image was chosen, and the average reflectance spectrum of each ROI was extracted to represent each sample by Envi4.6.

3 Multivariate chemometric methods

3.1 Principal component analysis

Principal component analysis (PCA) is usually used to reduce the dimensionality of the data while retaining most of the variance in the datasets^[6]. By several principal components (PCs), each sample can be represented by relatively few numbers instead of

thousands of variables^[6]. Generally, the first several PCs can reveal the most relevant information, and different scores of PCs can express the importance of each PC^[21].

3.2 Successive projection algorithm

The successive projection algorithm (SPA) is a forward selection method, which was used to select variable wavelengths from the full spectrum wavelengths. Select effective sets of variables for multivariate calibration, and it starts with one wavelength, then incorporates a new one at each iteration, until a specified number N of wavelengths is reached^[21].

3.3 Partial Least Squares

Partial Least Squares (PLS) is one method used for constructing predictive models. As one linear discrimination method, the PLS algorithm is commonly used in spectral data analysis and can effectively eliminate the co-linearity problem of spectral data^[8].

3.4 Least squares support vector machines

Least squares support vector machines (LS-SVM) methodology, an optimized version of the standard SVM, is one of the supervised learning methods (classes or composition of the samples in the data matrix is involved)^[21-23]. It has a wide application for pattern recognition and function estimation.

3.5 Software tools

The spectra data extraction was conducted on Envi4.6, PCA and PLS models were carried out on Unscrambler X10.1 (CAMO AS, Oslo, Norway), LS-SVM models were built using LS-SVM v1.8 toolbox running on MATLAB R2011a(The MathWorks and all graphs were designed by Origin8 SR0 (Origin Lab Corporation, Northampton, MA, USA).

4 Results and discussion

4.1 Spectral profiles

The average spectral curves of damaged wheat and healthy wheat covering the range of 865-1711 nm were illustrated in Figure 3. It could be observed that the general trends of spectral reflectance curves of the two groups showed similar profiles. From the range of 865-1130 nm, the spectral reflectance of the damaged wheat was higher than healthy wheat, and the trend of curves was fluctuant; From the range of 1130-1160 nm, the spectral reflectance of the damaged wheat showed similar spectral curves with healthy wheat and slight differences of reflectance value; From the range of 1160-1711 nm, the spectral reflectance of the damaged wheat was lower than healthy wheat, which was obvious as the wavelength increased.

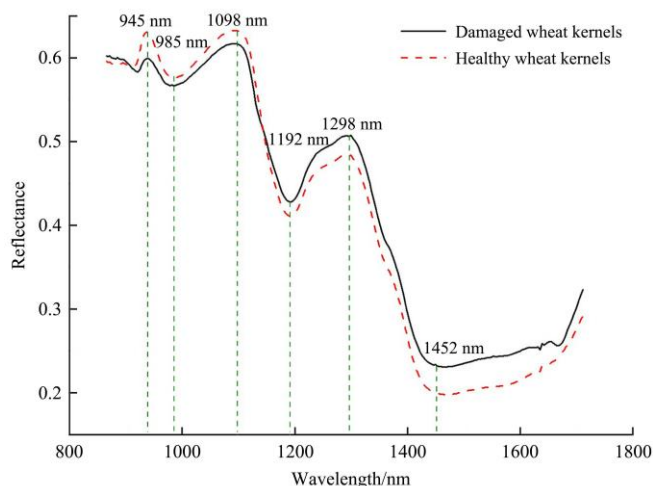


Figure 3 Mean spectral reflectance curves of wheat kernels

The locations of the reflective valley and reflection peak were almost the same. There were obvious reflective valleys around 985 nm, 1192 nm and 1452 nm. There were obvious reflection peaks at 945 nm, 1098 nm, and 1298 nm. They were caused by frequency doubling or frequency absorption of C-H, N-H, O-H and other functional groups. In detail, the small valley around 960-980 nm, corresponding to second- and first-overtone O-H stretching^[8], was associated with water strong absorption. The peaks and valleys around 1100 nm to 1300 nm, corresponding to second- and first-overtone C-H stretching, and the significance of wavelengths in this region was related to starch molecules in the wheat kernels. Meanwhile, the valley around 1450 nm was assigned to the C-H combination band^[8].

4.2 Principal component analysis

PCA was used for qualitative analysis of discriminating damaged wheat kernels from the healthy wheat kernels. PCA was performed on the preprocessed spectral of the two classes. And the corresponding score plots based on PC1, PC2, PC3 were shown in Figure 4. In the group of damaged wheat kernels and healthy wheat kernels, PC1, PC2 and PC3 explained 75%, 19% and 2% of the variations among samples, respectively. From Figure 4, there was obvious differentiation between the two groups of wheat kernels samples, and still, slightly overlaps were observed. To discriminate damaged wheat kernels from healthy ones correctly, the corresponding X-loadings of PC1, PC2, and PC3 would be analyzed, and the effective wavelengths would be selected. And discriminant models needed to be investigated for classifying the two classes.

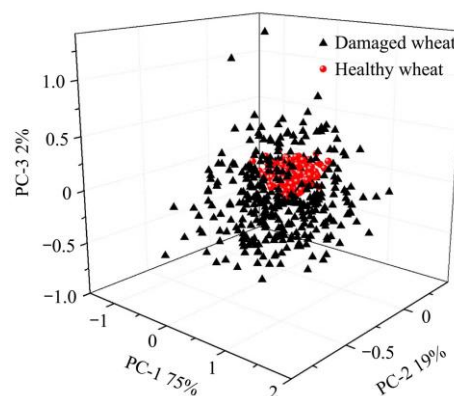


Figure 4 3-D cluster plots based on PC1, PC2, PC3

4.3 Effective wavelengths selection

Effective wavelengths selection aims to select only a few wavelengths that carry most of the useful information with minimum replacing full spectra. In this study, the loading values of PCA and SPA methods were used to select effective wavelengths to reduce data dimensionality.

4.3.1 X-loadings method

Based on PCA, X-loadings of the first 3 PCs were applied to identify effective wavelengths. Figure 5 showed the loading plots and the corresponding labeled effective wavelengths of the first 3 PCs from the PCA. Generally, peaks and valleys in loading plots, which revealed the relatively high absolute loading values, were identified as the effective wavelengths for discriminating damaged wheat kernels from healthy wheat kernels. Figure 5 showed loading had big valleys or peaks at wavelengths 937 nm, 965 nm, 1129 nm, 1176 nm, 1217 nm, 1340 nm, 1411 nm and 1677 nm. The corresponding identified effective wavelengths were labeled in different curves in Figure 5 with the arrow pointing and were shown in Table 1.

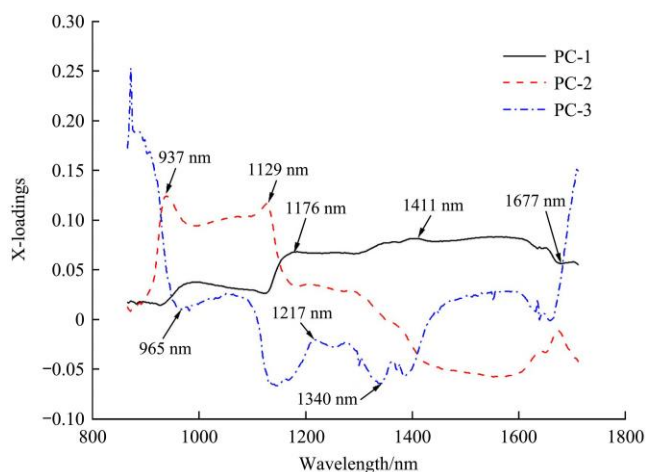


Figure 5 X-loadings plots of the PC1, PC2 and PC3 of PCA

Table 1 Effective wavelength selected by two methods

Method	Number	Wavelength/nm
X-loading	8	937, 965, 1129, 1176, 1217, 1340, 1411, 1677
SPA	8	935, 961, 1130, 1207, 1300, 1370, 1449, 1672

4.3.2 SPA method

SPA was used to select effective wavelengths of determining damaged wheat kernels from the healthy ones and the effective wavelengths selected from SPA were 935 nm, 961 nm, 1130 nm, 1207 nm, 1300 nm, 1370 nm, 1449 nm and 1672 nm. All these effective wavelengths are shown in Table 1.

4.4 Classification models

PLS and LS-SVM methodology were employed to establish models based on the calibration dataset for discriminating the damaged wheat kernels from the healthy wheat kernels. To build the PLS and LS-SVM supervised classification models, the damaged wheat kernels and healthy wheat kernels were firstly assigned values of 1 and 2. Then, samples of each category were divided into a calibration set and a prediction set with a ratio of 3:1, resulting in 270 samples and 90 samples in the calibration and the prediction set, respectively.

In detail, the Gaussian RBF kernel function was chosen, and the parameters of γ and σ^2 were optimized by a grid search procedure and 10-fold cross-validation. The performances of LS-SVM models were evaluated by classification accuracy. PLS models and LS-SVM models were established on full spectra, effective wavelength selected by X-loadings and SPA respectively. The classification results of PLS and LS-SVM models were shown in Table 2 and Table 3. From Table 2, It was found that the accuracy of discriminating healthy wheat kernels was all 100%, whatever on PLS models or LS-SVM models. All the healthy wheat kernels were successfully identified without samples misclassified. While, as to the damaged wheat kernels, several samples were misclassified as healthy wheat kernels with PLS models, there was no misclassification on LS-SVM models.

Overall, as shown in Table 3, LS-SVM models using full spectral, the effective wavelengths selected from X-loading and SPA all got satisfactory classification results, 100% respectively. And PLS models on corresponding datasets obtained satisfactory accuracy too, 98.9%, 98.9% and 97.2% respectively. So LS-SVM models on effective wavelength obtained a better performance, which could be used for discriminating damaged wheat samples from the healthy ones, revealing the advantage of hyperspectral imaging technology.

Table 2 Classification results of wheat

Classification model	Wavelength Selection	variable	Prediction		Accuracy /%
			1	2	
PLS	Full spectral	1	88	2	97.8
		2	0	90	100
	X-loading	1	88	2	97.8
		2	0	90	100
	SPA	1	87	3	96.7
		2	0	90	100
LS-SVM	Full spectral	1	90	0	100
		2	0	90	100
	X-loading	1	90	0	100
		2	0	90	100
	SPA	1	90	0	100
		2	0	90	100

Table 3 Total classification accuracy on PLSR and LS-SVM

Wavelength selection	PLS/%	LS-SVM/%
Full spectral	98.9	100
X-loading	98.9	100
SPA	97.2	100

5 Conclusions

In this study, hyperspectral imaging was applied to discriminate against the damaged wheat kernels from the healthy wheat kernels. Effective spectral wavelength was obtained by X-loadings and SPA methods. Then PLS models and LS-SVM models were built on full spectral, and effective spectral datasets, respectively. All the models got satisfactory results, with classification accuracy over 97%. And the LS-SVM models performed best with classification accuracy being 100%. However, several damaged wheat kernels samples were misclassified as healthy samples with PLS models. The results of this study indicated that hyperspectral imaging has the potential of discriminating damaged kernels in wheat samples. In further, more samples with different proportion of damaged wheat kernels were needed for robust discrimination of wheat kernels grading. Additionally, the development of the portable detectors on the damaged wheat kernels would be beneficial for wheat kernels grading or post-harvest quality processing of other grains.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 31671632; No. 31701325), Green Farming and Mechanical Innovation Team of Fruit Harvesting under Soil.

[References]

- [1] AQSIQ, SAC. National Standard of the People's Republic of China: Wheat, GB1351-2008, 2008; 6p.
- [2] AQSIQ, SAC. National Standard of the People's Republic of China: Grain and oil inspection of food, oil impurities, damaged kernels inspection, GB/T5494-2008, 2008; 6p.
- [3] Kimuli D, Wang W, Lawrence K C, Yoon S C, Ni X, Heitschmidt G W. Utilisation of visible/near-infrared hyperspectral images to classify aflatoxin B 1 contaminated maize kernels. *Biosystems Engineering*, 2018; 166: 150-160.
- [4] Caporaso N, Whitworth M B, Fisk I D. Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chemistry*, 2018; 240: 32-42.

- [5] Senthilkumar T, Jayas D S, White N D G, Fields P G, Gräfenhan T. Detection of Ochratoxin A contamination in stored wheat using near-infrared hyperspectral imaging. *Infrared Physics & Technology*, 2017; 81: 228–235.
- [6] Lancelot E, Bertrand D, Hanafi M, Jaillais B. Near-infrared hyperspectral imaging for following imbibition of single wheat kernel sections. *Vibrational Spectroscopy*, 2017; 9: 46–53.
- [7] Senthilkumar T, Jayas D S, White N D G, Fields P G, Gräfenhan T. Detection of fungal infection and Ochratoxin A contamination in stored wheat using near-infrared hyperspectral imaging. *Journal of Stored Products Research*, 2016; 65: 30–39.
- [8] Chu X L. Near-infrared spectroscopy analytical technology practical handbook. Beijing: Mechanical Industry Press, 2016; 493p.
- [9] Williams P J, Kucheryavskiy S. Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*, 2016; 209: 131–138.
- [10] Barreto A, Cruz-Tirado J P, Siche R, Quevedo R. Determination of starch content in adulterated fresh cheese using hyperspectral imaging. *Food Bioscience*, 2018; 21: 14–19.
- [11] Su W H, Sun D W. Evaluation of spectral imaging for inspection of adulterants in terms of common wheat flour, cassava flour and corn flour in organic Avatar wheat (*Triticum* spp.) flour. *Journal of Food Engineering*, 2017; 200: 59–69.
- [12] Esteki M, Vander Heyden Y, Farajmand B, Kolahderazi Y. Qualitative and quantitative analysis of peanut adulteration in almond powder samples using multi-elemental fingerprinting combined with multivariate data analysis methods. *Food Control*, 2017; 82: 31–41.
- [13] Lohumi S, Lee S, Lee H, Cho B K. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science & Technology*, 2015; 46(1): 85–98.
- [14] Verdú S, Vázquez F, Grau R, Ivorra E, Sánchez A. J, Barat J M. Detection of adulterations with different grains in wheat products based on the hyperspectral image technique: The specific cases of flour and bread. *Food Control*, 2016; 62: 373–380.
- [15] Guo D, Zhu Q, Huang M, Guo Y, Qin J. Model updating for the classification of different varieties of maize seeds from different years by hyperspectral imaging coupled with a pre-labeling method. *Computers and Electronics in Agriculture*, 2017; 142: 1–8.
- [16] Singh C B, Jayas D S, Paliwal J, White N D G. Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging. *Journal of Stored Products Research*, 2009; 45(3): 151–158.
- [17] Singh C B, Jayas D S, Paliwal J, White N D G. Identification of insect-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging. *Computers and Electronics in Agriculture*, 2010; 73(2): 118–125.
- [18] Dong J J, Wu J Z, Liu Q. Research on hyperspectral image detection method of wheat unsound kernel. *Journal of Electronic Measurement and Instrumentation*, 2017; 31(7): 1074–1080. (in Chinese)
- [19] Luo X, Jayas D S, Symonst S J. Identification of damaged kernels in wheat using a colour machine vision system. *Journal of Cereal Science*, 1999; 30: 49–59.
- [20] Narvankar D S, Singh C B, Jayas D S, White N D G. Assessment of soft X-ray imaging for detection of fungal infection in wheat. *Biosystems Engineering*, 2009; 103(1): 49–56.
- [21] Shao Y Y, Xuan G T, Hu Z C, Gao Z M, Liu L. Determination of the bruise degree for cherry using Vis-NIR reflection spectroscopy coupled with multivariate analysis. *PLoS One*, 2019; 14(9): 1–13.
- [22] Gao Z M, Zhao Y R, Khot L R, Hoheisel G A, Zhang Q. Optical sensing for early spring freeze related blueberry bud damage detection: Hyperspectral imaging for salient spectral wavelengths identification. *Computers and Electronics in Agriculture*, 2019; 167: 105025. doi: 10.1016/j.compag.2019.105025.
- [23] Gao Z M, Shao Y Y, Xuan G T, Wang Y X, Liu Y, Han X. Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. *Artificial Intelligence in Agriculture*, 2020; 4: 31–38.