

Rapid and non-destructive decay detection of Yali pears using hyperspectral imaging coupled with 2D correlation spectroscopy

Yufan Zhang^{1†}, Wenxiu Wang^{1,2†}, Fan Zhang¹, Qianyun Ma¹, Shuang Gao¹, Jie Wang¹, Jianfeng Sun^{1*}, Yuanyuan Liu²

(1. College of Food Science and Technology, Hebei Agricultural University, Baoding 071000, Hebei, China;

2. Agricultural Engineering Key Laboratory at Universities of Education Department of Xinjiang Uygur Autonomous Region, Tarim University, Alar 843300, Xinjiang, China)

Abstract: The black spot disease caused by *Alternaria alternata* on Yali pears is a great concern as it compromises their edible quality and commercial value. To realize rapid and non-destructive classification of this disease, hyperspectral imaging (HSI) technology was combined with two-dimensional correlation spectroscopy (2DCOS) analysis. A total of 150 pear samples at different decay grades were prepared. After obtaining the HSI images, the whole sample was demarcated as the region of interest, and the spectral information was extracted. Seven preprocessing methods were applied and compared to build the classification models. Thereafter, using the inoculation day as an external perturbation, 2DCOS was used to select the feature-related wavebands for black spot disease identification, and the result was compared to those obtained using competitive adaptive reweighting sampling and the successive projections algorithm. Results demonstrated that the simplified least squares support vector model based on 2DCOS-identified feature wavebands yielded the best performance with the identification accuracy, precision, sensitivity, and specificity of 97.30%, 94.60%, 96.16%, and 98.21%, respectively. Therefore, 2DCOS can effectively interpret the feature-related wavebands, and its combination with HSI is an effective tool to predict black spot disease on Yali pears.

Keywords: hyperspectral imaging technology, black spot disease, two-dimensional correlation spectroscopy, Yali pear

DOI: 10.25165/ijabe.20221505.7313

Citation: Zhang Y F, Wang W X, Zhang F, Ma Q Y, Gao S, Wang J, et al. Rapid and non-destructive decay detection of Yali pears using hyperspectral imaging coupled with 2D correlation spectroscopy. Int J Agric & Biol Eng, 2022; 15(5): 236–244.

1 Introduction

The Chinese pear ‘Yali’ (*Pyrus bretschneideri* Rehd) is a variety in the ‘White pear’ system, which is rich in vitamin C and minerals, such as calcium, phosphorus, and iron. It is flavorful and popular among consumers in China^[1]. However, during storage, Yali is prone to infection by *Alternaria alternata* and, consequently, the black spot disease. With fungal infections, brown spots appear locally on the surface of pears^[2]. Additionally, the physicochemical property value of the diseased pear changes, resulting in quality loss, which critically affects its commercial value. Thus, it is crucial to detect and classify the black spots in

pears. Presently, the conventional method for black spot disease identification and detection is recognition by the human eye and morphological identification of *Alternaria alternata*. However, these methods have the disadvantages of low efficiency, time-consuming, and laborious. Additionally, the inefficiency of the industrially prevalent method of simple visual inspection by trained personnel and the high degree of subjectivity in the inspection process are major concerns^[3]. Therefore, the development of a rapid nondestructive detection method is needed to identify black spots in Yali pear.

Hyperspectral imaging (HSI) combines near-infrared spectroscopy and two-dimensional imaging to obtain spatial and spectral information simultaneously^[4]. In recent years, HSI has been widely investigated as a promising technology for disease identification in fruits and vegetables, such as peach^[5], mandarin^[6], apple^[7], citrus^[8], potato^[9], and onion^[10]. All of these studies indicate that HSI is effective for the recognition of fruit and vegetable diseases.

However, the acquired hyperspectral images typically have high dimensional data and contain a large amount of redundant information. The collinearity and redundancy of images at contiguous wavelengths reduce the computational efficiency and deteriorate the performance of the model^[11]. To solve these problems, characteristic wavelength selection is generally implemented to select informative variables and remove uninformative and noisy ones while ensuring a shorter duration for computational modeling^[12]. Some of the methods for selecting characteristic wavelengths are competitive adaptive reweighted sampling algorithm (CARS), genetic algorithm, and successive

Received date: 2021-12-31 **Accepted date:** 2022-07-22

Biographies: **Yufan Zhang**, Master candidate, research interest: rapid nondestructive testing of food spectra, Email: 13731233684@163.com; **Wenxiu Wang**, PhD, Associate Professor, research interest: rapid nondestructive testing of food spectra, Email: godlovexiu@163.com; **Fan Zhang**, Master candidate, research interest: rapid nondestructive testing of food spectra, Email: zfl8730278721@163.com; **Qianyun Ma**, PhD, Associate Professor, research interest: extraction, analysis and utilization of natural active ingredients, Email: maqianyun@126.com; **Shuang Gao**, Master, Lecturer, research interest: food science, Email: 305147043@qq.com; **Jie Wang**, PhD, Professor, research interest: marine aquatic products processing, Email: wj591020@163.com; **Yuanyuan Liu**, PhD, Associate Professor, research interest: research and equipment research on quality and safety of agricultural products, Email: lyylvjin@163.com.

† These authors contributed equally to this study.

***Corresponding author:** **Jianfeng Sun**, PhD, Professor, research interest: food processing, safety and equipment. College of Food Science and Technology, Hebei Agricultural University, Baoding 071001, Hebei, China. Tel: +86 312-7528199, Email: causunjf@126.com.

projection algorithm (SPA). Although they are commonly used, the selection criteria of most of these methods are not driven by real relevance to spectral features; instead, they use a searching algorithm based on prediction accuracy from a statistical perspective. Further, the selected variables representing a functional group of the analytes or a property of interest were poorly interpreted^[13]. In comparison with traditional one-dimensional (1D) spectra, two-dimensional correlation spectroscopy (2DCOS) can enhance spectral resolution by spreading the original data over the second dimension, and emphasize the features not readily observable in 1D spectra^[14]. It can discern small spectral changes and interpret them at the molecular level. The 2D correlation spectra provide chemical fingerprints that aid in interpreting chemometric models. This facilitates the identification of the chemical components present in the sample. Currently, this technology is used to identify adulterated rice^[15] and edible oils^[16]. However, the combination of 2DCOS and HSI to identify black spots in the Yali pear has rarely been reported.

This study attempted to combine HSI and 2DCOS to detect black spot disease in Yali pears. The main objectives are as follows: 1) acquisition of hyperspectral images of Yali pears with and without black spots; 2) region of interest (ROI) segmentation from hyperspectral images and extraction of average spectra; 3) pretreatment of original spectra; 4) selection of characteristic bands using 2DCOS, CARS, and SPA; 5) use of the least-squares support-vector machine (LS-SVM) for modeling based on the full band spectra and characteristic bands. In this study, an efficient method was expected to be developed to interpret HSI spectra of diseased pears and identify the most effective variables in a rapid and accurate manner.

2 Materials and methods

2.1 Sample preparation

Fresh Yali pear fruits were purchased from a local fruit market in Baoding, Hebei, China, and were immediately transported to the laboratory of Hebei Agricultural University, where were stored at 4°C prior to use. A total of 151 pears with comparable maturity, shape, and weight and free from apparent defects, physical injuries, or disease infection were selected. They were cleaned with detergent, surface-sterilized with 70% ethanol to eliminate other microbes, rinsed thrice with distilled water, and finally air-dried at room temperature (25°C).

Six Yali pears with natural black spot disease were selected as the source of pathogenic bacteria. After being purchased from a local farmers' market in Baoding, Hebei, China, the diseased pear was wiped and disinfected with alcohol cotton balls, and the pericarp was removed with a sterile knife. Then, the diseased pulp was placed on a prepared potato dextrose agar (PDA) plate and stored in a 25°C incubator. After 3 d, new mycelium grew around the pulp, and they were picked by an inoculation ring for purification on another PDA plate. After 7 d of culture at 25°C, the pathogen spores on the PDA plate were scraped off with a triangular glass rod, suspended in sterile distilled water containing 0.05% Tween-80, and placed into a 50 mL triangular flask. The flask was then shaken for 15 s and filtered with four layers of gauze to make the sporozoan suspension. The concentration of conidial suspension was adjusted to about 10⁶ spores conidia per milliliter with the aid of a hemocytometer^[17].

A 20 μL of *Alternaria alternata* conidial suspension was injected 3 cm deep into the fruit's equator with a sterile syringe

perpendicular, and the fruits were incubated at 24°C-25°C, 80% RH. Seven pear fruits were inoculated daily, and the experiment lasted for 18 d. The lesion diameters of the disease spots on each fruit were recorded during cultivation, and samples were classified according to the grading standard shown in Table 1^[18]. All samples were classified into four grades: healthy samples (Grade0, G0), light decayed samples (Grade1, G1), medium decayed samples (Grade2, G2), and severely decayed samples (Grade3, G3). After the elimination of one abnormal decayed sample, a total of 25 fresh and 125 inoculated pear fruit were acquired for subsequent analysis.

Table 1 Grading standard for Yali pears used in this study

Category	Grading standard	Number
Healthy (G0)	No inoculation, no disease spots	25
Light decayed (G1)	Percentage of spot area ≤ 20%	55
Medium decayed (G2)	20% < Percentage of spot area ≤ 60%	59
Severely decayed (G3)	Percentage of spot area > 60%	11

2.2 Hyperspectral imaging system

Images of pear fruit were acquired using a pushbroom HSI system in reflectance mode. The system mainly consisted of four components: a charge-coupled device camera (SPECIM FX 10, Specim company, Finland) with a resolution of 1024 pixels (pixel size is 8×8 μm) in the spatial dimension and 224 bands in the spectral dimension, a hyperspectral imaging workstation (SisuCHEMA, Specim company, Finland) with the spectral range of 398-1004 nm, two halogen lamps (JCR, 15 V, 150 W, BAU, Japan) and a computer with a hyperspectral image analysis software. The spectral resolution is 5.5 nm and the imaging speed of full-band acquisition is 330 fps.

The measured sample was placed on the mobile platform, and the lesion area was manually placed upward to facilitate image acquisition. To prevent image over-saturation, the speed of the moving platform, the camera exposure time, and the collection distance need to be set in advance. After repeated tests, the aforementioned three parameters were set as 9.74 mm/s, 8.50 ms, and 500 mm, respectively. The images acquired under this condition are clear and full. The camera records the spectral information reflected by the sample in the form of line scan, and a group of hyperspectral information of size (x, y, λ) can be obtained, where x and y are spatial dimensions (the number of rows and columns in pixels), and λ is the number of spectral bands.

Due to the dark current and different lenses' sensitivity to light, the captured images are not consistent with the real situation. Therefore, it is necessary to correct the data through black-and-white reference^[19]. A Teflon white board (Spectraon SRT-99-100, Labsphere Inc., North Sutton, NH, USA) was used to obtain the white image. The dark image was acquired by covering the lens with an opaque cap. The corrected relative image was calculated according to the formula as follows:

$$I = \frac{I_o - I_d}{I_w - I_d} \times 100\% \quad (1)$$

where, *I* is the corrected spectral image; *I_o* is the original spectral image; *I_w* is the white image; *I_d* is the dark image.

2.3 Spectral data acquisition

Spectral extraction of hyperspectral images was performed using the Environment for Visualizing Images software (ENVI 5.1, Research System Inc., Boulder, MT, USA). The selection of the region of interest (ROI) greatly affects the accuracy of the model. In order to contain as much information as possible, all the pixels of the pear were selected as ROI. The pear and background were

distinguished by the threshold segmentation method, which involves setting a gray value in advance, retaining the region greater than this value, and removing other regions. In this study, three points were selected at the junction of the samples and black background, and the average grayscale value was considered as the threshold value. Then the average spectrum of all pixel points was calculated by ENVI and considered as the spectral information of the sample. As the collected spectral images were clear in all bands, the whole spectral range with 224 bands was retained for the analysis.

2.4 Spectral pretreatment

To reduce or eliminate the influence of spectral drift or migration on the prediction ability of the model, the original spectra are treated by some pretreatment methods^[20]. In this study, the original spectral data underwent: 1) scattering correction (standard normal variate (SNV) correction), which was used to eliminate the scattering effects caused by the particle size or an uneven distribution within the spectrum^[21,22]; 2) mathematical pretreatments (first derivative (FD) treatment and second derivative (SD) treatment), which can eliminate baseline drift and baseline rotation^[23]; 3) combinations and different orders of use thereof (SNV-FD, SNV-SD, FD-SNV, SD-SNV). The spectral pretreatment was carried out in PLS_Toolbox 7.5 (Eigenvector Research Inc., Manson, WA, USA).

2.5 Effective variables selection

It is important to conduct wavelength selection to reduce the redundant information in the hyperspectral data for further designing a robust model^[24]. In this study, 2DCOS was used for identifying spectra features related to fruit decay from the full spectral range. Meanwhile, the commonly used methods CARS and SPA were also employed and compared with 2DCOS.

2DCOS was carried out using the 2DCOS software freely available on the website designed by Tao Zhou (<http://muchong.com/t-9696009-1>). 2DCOS extends the spectral information from one dimension to two dimensions, improves the resolution of the data, and makes it easier to discover the hidden information^[25]. It can help deeply understand the spectral changes affected by external perturbation at the molecular level. In this study, inoculation day was used as the external perturbation to select a series of dynamic spectra. Thus 19 samples were selected, including 18 samples with different decay grades and 1 fresh sample. Before the 2D-COS analysis, SD was used to process the spectra as it can remove some high-frequency noise and interference between components.

CARS is a fast and effective wavelength selection method. In general, each sampling run consists of four main steps. Firstly, Monte Carlo method was used to select N subsets (each subset contains 80%-90% of the calibration set) from the calibration set, and the calibration model was established with each subset. Secondly, the exponential decay function (EDF) was used to remove little or no information in the whole spectrum gradually and effectively. Thirdly, adaptive weighted sampling (ARS) simulates the "survival of the fittest" principle by selecting heavier wavelengths. Finally, the wavelengths with the smallest root mean square error of cross-verification (RMSECV) were the optimal wavelengths. The SPA algorithm is also widely used in the wavelength selection of hyperspectral data. Vector projection analysis was used to find out the minimum redundant information variable group in the spectral information, so as to minimize the collinearity between variables and reduce the number of modeling variables. In SPA, multiple models based on different subsets

were established and the root mean square error (RMSE) of the models was calculated. The optimal wavelengths can be determined from a subset of the minimum RMSE.

2.6 Model establishment

All samples were divided into a calibration set and a prediction set by kennard-stone (K-S) algorithm according to a ratio of 3:1. Then there were 113 samples in the calibration set and 37 samples in the prediction set. After the sample partition, the spectral data in the calibration set was used as the input of LS-SVM, and the identification models were verified with the samples in the prediction set. LS-SVM (Suykens, <http://www.esat.kuleuven.ac.be/sista/members/suykens.html>) is an improved version of the support vector machine and is used to analyze data and pattern recognition as well as classification and regression analyses^[26]. In addition, LS-SVM runs significantly faster than other versions of support vector machines due to its least-squares approach. Compare with methods like convolutional neural network, the advantages of LS-SVM include short training time, small training sample, strong generalization ability, and high accuracy. The aforementioned data analyses were all implemented in MATLAB2014a (The Mathworks Inc., Natick, MA, USA).

2.7 Model evaluation

Average recognition accuracy, precision, sensitivity, and specificity are the general measurement metrics to evaluate the performance of the classification model. The performance of the established LS-SVM model was evaluated by calculating the accuracy, precision, sensitivity, and specificity according to Equations (2)-(5)^[27].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (5)$$

where, TP represents the number of individuals that are actually in category C (the target category) and are correctly classified by the classifier; FP represents the number of individuals that do not belong to category C but are misclassified into category C; TN represents the number of individuals that do not belong to category C and are correctly classified by the classifier; FN represents the number of individuals that belong to category C but are misclassified by the classifier. The hyperspectral image analysis structure is shown in Figure 1.

3 Results and analysis

3.1 Sample analysis

Following the inoculation of pears with a fungal suspension, the wound on the fruit surface was barely visible in the early stage of infection. Then, necrosis of the epidermal tissue at the wound resulted in the formation of a little black spot that gradually expanded as the grade of corruption increased. The spot was characterized by a grayish-brown round shape and became more apparent and darker as time progresses. This may be because, as the disease progresses, pathogens proliferated into healthy areas around it and began to produce symptoms. In severely decayed fruits, the diseased areas spread further to most of the fruits, and significant fruit spot was observed. Some liquid precipitation was

observed in the lesion area, which was probably caused by the proliferation of pathogenic microorganisms; this phenomenon

destroys the cell structure and leads to the exudation of cell fluid. Additionally, a few brown spots were formed around the lesion area.

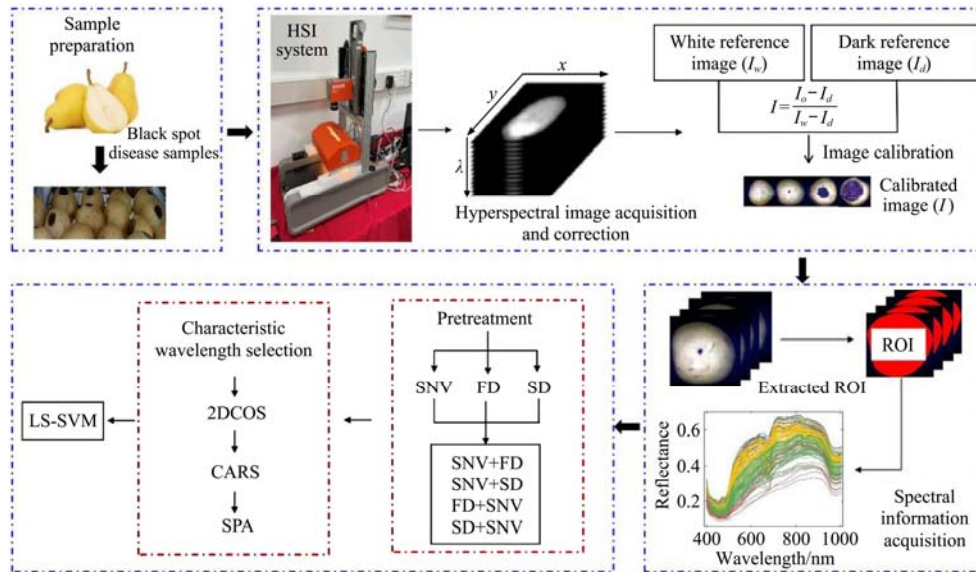


Figure 1 Hyperspectral image analysis structure

3.2 Spectral characteristics

The original spectra of all samples are shown in Figure 2a, where the blue, yellow, green, and red lines represent the fresh, light-decayed, medium-decayed, and severely-decayed samples, respectively. The state of samples with different decayed degrees is shown in the lower-right corner of Figure 2a. It can be observed that the trends of the spectra of pear samples with and without different corruption grades were consistent; however, the spectral reflectance intensity at different wavelengths varied, indicating that different samples comprised similar substances, but different quantities of these substances. The spectral reflectance of healthy pears was significantly higher than that of decayed pears and generally decreased continuously with an increase in corruption grades. The decrease in spectral reflectance represents the change in the characteristics of the fruit, including hidden information about different components such as sucrose, fructose, and glucose. Three pronounced absorption peaks were observed at 485 nm, 650 nm, and 980 nm. The peaks at 485 nm and 650 nm correspond to the absorption peaks of carotenoids and chlorophyll, and the absorption peak near 980 nm is related to the first overtone of O-H^[28,29]. In addition, there are small absorption peaks at 740 nm and 780 nm, which represent the absorption peaks of water and cellulose, respectively^[30]. The inoculated fruit showed reduced chlorophyll and water content so the two absorption peaks of their spectra decreased. The absorption peak at 650 nm was different for some samples of the same grade. The reason might be that there are inevitable differences among pear samples, such as inconsistent original chlorophyll content. Moreover, the absorption peak at 680 nm gradually disappeared with the increase in decayed grades owing to the decomposition of chlorophyll during the corruption process due to its instability.

In general, owing to the complex relationship between different wavelengths and functional groups in the entire spectral range, there is no evident absorption peak that corresponds directly to the disease grade in the reflectance map. Therefore, it is necessary to analyze the disease grades of the samples using modeling.

3.3 Spectral pretreatment

Figures 2b-2h show the ability of several pretreatment methods to correct the raw spectra. After preprocessing, the spectral data

retained the absorption characteristics of the raw spectra and improved to a certain degree. The interference of the baseline and background is eliminated when the spectra are processed by FD and SD, and the characteristic peaks in the original spectra are highlighted. This implies that the FD and SD pretreatments strengthen the small characteristic peaks hidden in the wider absorption frequency band in the original spectra. Moreover, the corresponding differences in the absorbance of local positions such as extreme and inflection points are also strengthened. After SNV treatment, the spectra were consistent with the original spectra; additionally, the spectra became concentrated between 400 and 640 nm (Figure 2d), indicating that the effect of scattering was weakened. Generally, SNV, FD, and SD have successfully reduced spectral dispersion and corrected the baseline shift.

As individual spectral pretreatment may not offer ideal prediction results owing to the complexity of decayed pear fruit samples, the combinations of different pretreatment methods were also employed to process the original spectra. As shown in Figures 2e and 2g, the trends of these two spectra are consistent, but the variations between samples in the FD-SNV spectra are greater than those in SNV-FD. Similarly, Figures 2f and 2h show the spectra pretreated by SNV-SD and SD-SNV. The trends of these two spectra are consistent, but the differences between samples in the SNV-SD spectra are greater than those in SD-SNV. Direct observation of spectral signal characteristics cannot indicate the optimal pretreatment method; this is due to the fact that the selection factors are subjective in nature. This leads to incorrect results. Therefore, evaluation should be conducted according to the prediction results of further modeling.

3.4 Model based on the entire spectra

Based on the full spectra preprocessed by FD, SD, SNV, SNV-FD, SNV-SD, FD-SNV, SD-SNV, and untreated spectra, classification models were developed. Table 2 presents the model performance for the prediction set in terms of overall accuracy, precision, sensitivity, and specificity. Generally, model prediction accuracy increases with the proximity of the four indicator values to 1. Compared to the original spectra, the prediction performance of the LS-SVM models developed based on the preprocessed spectral data, was satisfactory. The SNV-SD model

yielded the highest accuracy (95.95%), with precision, sensitivity, and specificity ranging from 76.73%, 82.33%, 94.08% to 88.35%, 94.60%, and 97.47%, respectively, implying that the SNV-SD

method was most suitable for the classification of black spot decayed grades. Additionally, suitable spectral preprocessing improves the identification accuracy of the models.

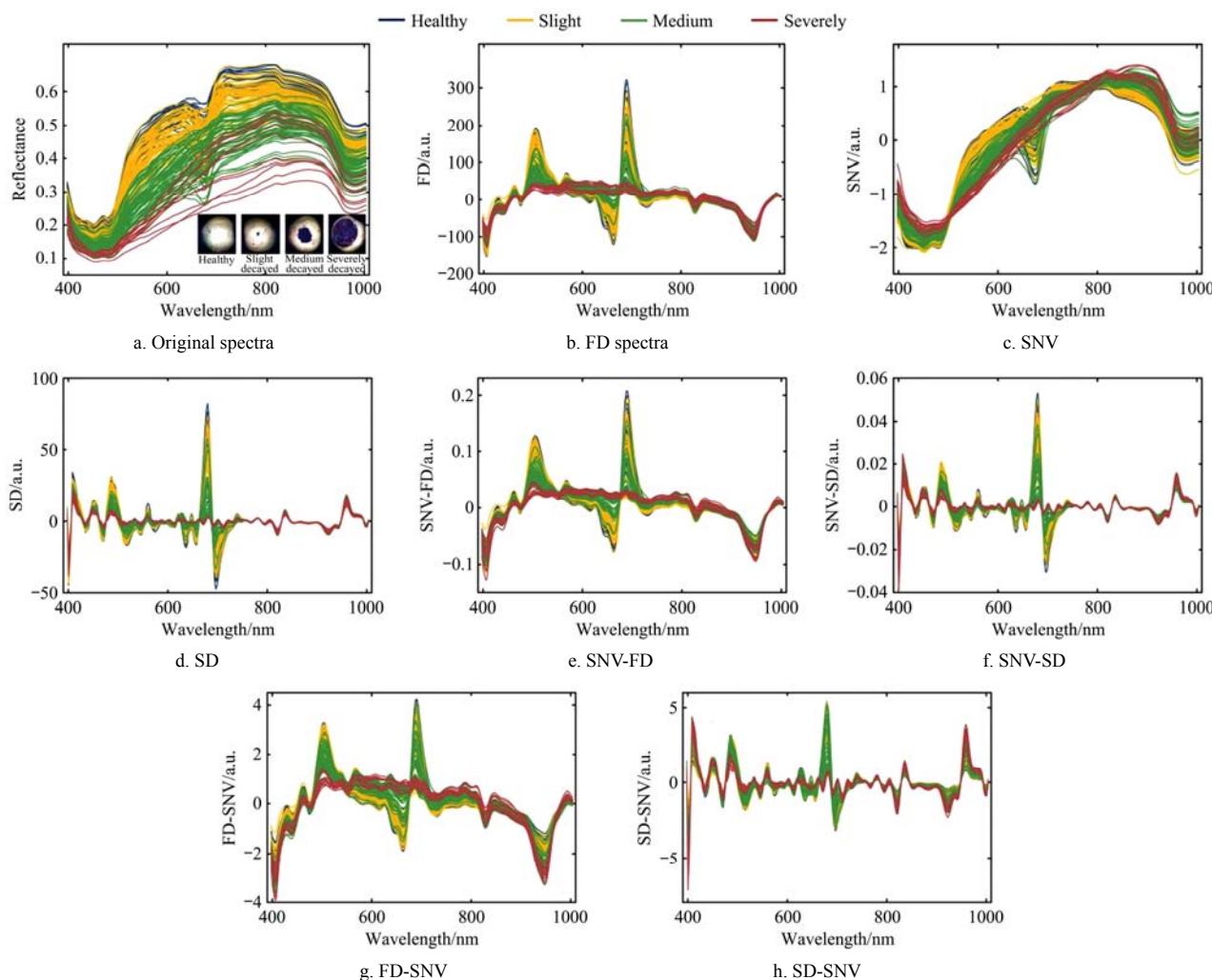


Figure 2 Original and preprocessed spectra

Table 2 Results of LS-SVM models based on the original and preprocessed spectra

Pretreatment	Accuracy/%	Precision/%	Sensitivity/%	Specificity/%
Original	90.54	76.73	82.33	94.08
FD	93.24	82.59	90.77	95.54
SD	93.24	83.97	88.76	95.32
SNV	94.59	85.94	93.04	96.38
SNV-FD	94.59	85.94	93.04	96.38
SNV-SD	95.95	88.35	94.60	97.47
FD-SNV	94.59	88.39	93.75	96.38
SD-SNV	94.59	88.39	91.74	96.38

Note: FD: First derivative; SD: Second derivative; SNV: Standard normal variate.

To further explore the detailed classification performance for each grade, the confusion matrices of the prediction results using the LS-SVM models established by different spectra are shown in Figure 3. For G0 pear samples, the models based on pretreated spectra presented better results than those based on the original spectra. For G1 pear samples, better results were obtained for the models based on spectra pretreated by FD-SNV and SD-SNV. For G2 pear samples, SD and SNV-SD pretreated spectra models exhibited better results. However, for G3 pear samples, pretreatment by SD and SNV-SD presented results identical to the original spectra, which were all satisfactory results as all the

samples were classified correctly. In terms of overall recognition ability, the SNV-SD model only misjudged three samples, which was the least among all models. Therefore, the SNV-SD model was selected for further optimization.

3.5 Model based on characteristic wavelengths

The analysis of the models with different pretreatments shows the excellent prediction performance of the SNV-SD model. Therefore, all further characteristic wavelength selections were conducted based on the SNV-SD-treated spectra using 2DCOS, CARS, and SPA methods.

3.5.1 Model based on characteristic wavelengths using 2DCOS

To obtain more detailed information, the spectra were divided into four sub-bands and two-dimensional correlation analysis was implemented. Figures 4a-4d show the synchronous two-dimensional correlation spectra in the 398-1004 nm interval. It can be seen that the overlapping peaks at the original spectra become clearer. According to Noda's two-dimensional spectral theory, the synchronous two-dimensional correlation spectra are symmetric about the main diagonal. The peak on the diagonal is the automatic peak, which is always positive. Its intensity reflects the strength of the system to be tested with a change in external perturbation, and the number of circles reflects its automatic peak strength. The greater the number of circles, the stronger the automatic peak will be.

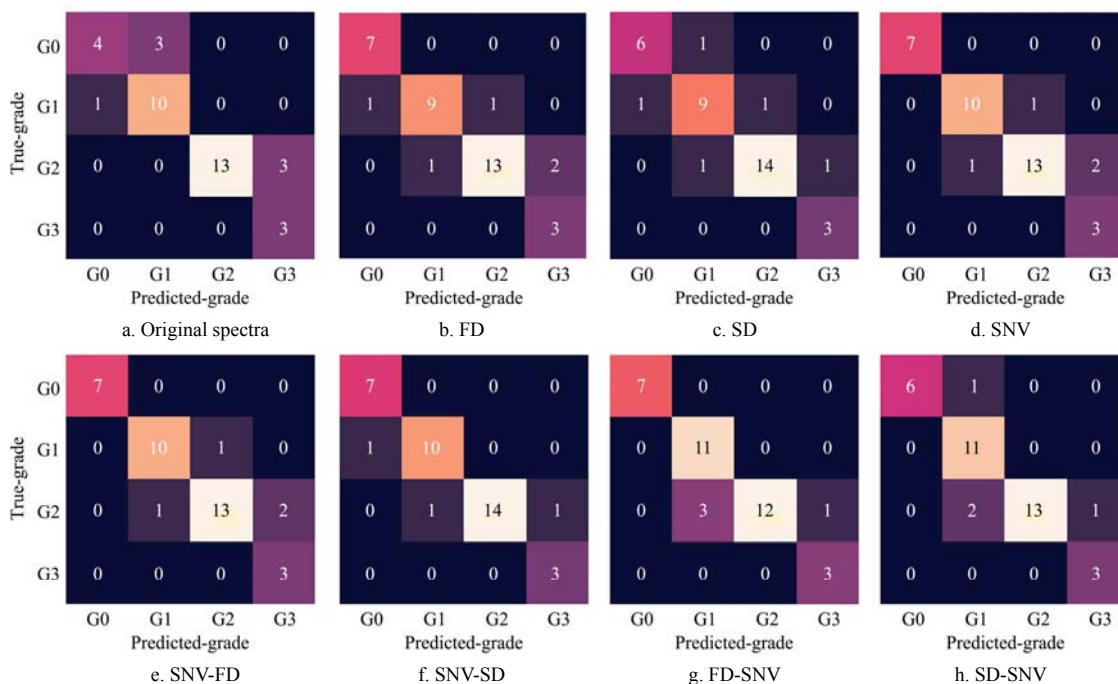


Figure 3 Confusion matrixes of the prediction results for each grade sample using the LS-SVM models based on original spectra and preprocessed spectra using different methods

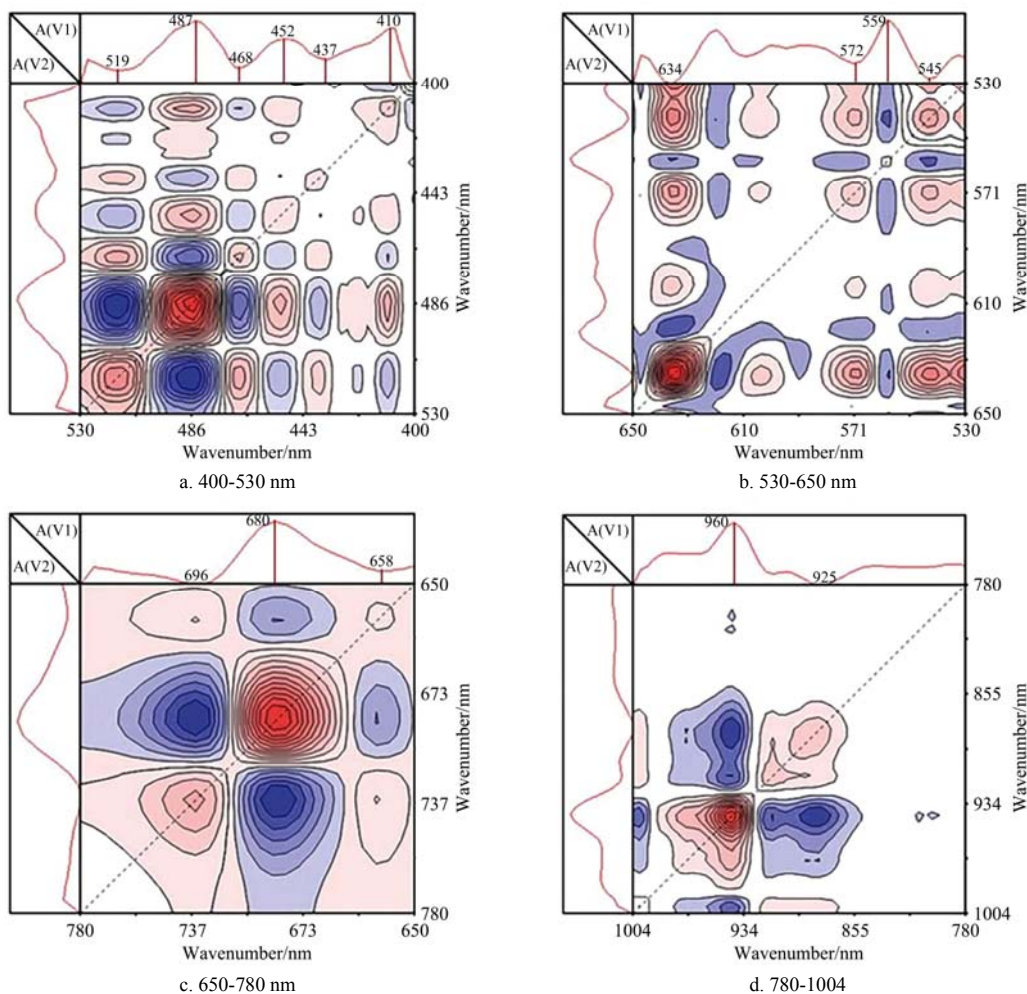


Figure 4 Two-dimensional correlation spectra at different bands

There are six autocorrelation peaks in Figure 4a, located at 410, 437, 452, 468, 487, and 519 nm respectively; Figure 4b has four autocorrelation peaks located at 545, 559, 572, and 634 nm respectively; Figure 4c has three autocorrelation peaks located at

658, 680, and 696 nm respectively; and Figure 4d has two autocorrelation peaks located at 925 and 960 nm, respectively. The (487, 487) nm and (680, 680) nm peaks are both related to chlorophyll absorption from the skin and flesh of pear fruit, while

(696, 696) nm is related to cellulose absorption. The 925 nm spectra were ascribed to water, starch, and cellulose, and those at 960 nm to water and sugar content. With progressive decay, the water and sugar content in the pears decreased continuously. These peaks can effectively represent the variation caused by an external disturbance in the sample, such as time. Therefore, these 15 wavelengths were selected as the characteristic wavelengths to develop the model.

Taking the preprocessed spectra by SNV-SD at these wavelengths as input and different decayed grades as output, the simplified LS-SVM model was established again. The performance of the model in terms of each grade is presented in the

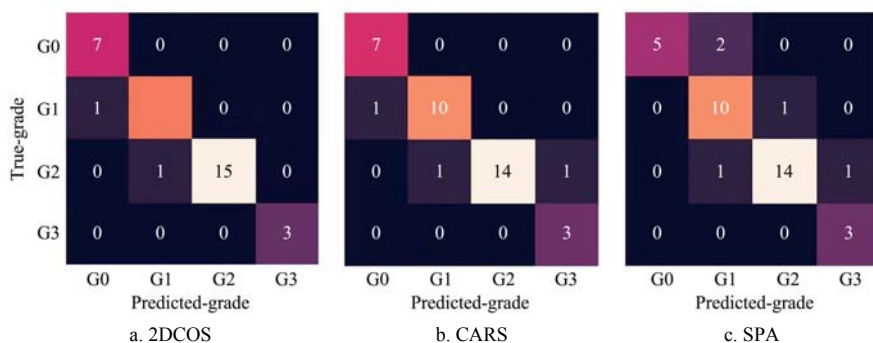


Figure 5 Confusion matrix of the classification result of the model based on selected characteristic wavelengths

3.5.2 Model based on characteristic wavelengths using CARS

The CARS algorithm was used to select effective wavelengths based on full spectral data by SNV-SD treatment. For the CARS algorithm, the number of Monte Carlo Sampling (MCS) runs was set to 1000, and the number of selected wavelengths was determined by 10-fold cross-validation. The process of CARS is shown in Figure 6, where Figure 6a is the trend of the screening variables, which can be divided into two stages, i.e., rapid and fine selection. As shown in Figure 6a, the sampling variables decreased as the number of runs increased. In the beginning, owing to the use of the EDF, the number of selected wavelengths drops sharply and thereafter becomes relatively stable. Figure 6b shows the trend diagram for RMSECV. Because of the elimination of uninformative wavelengths, there is a reduction in RMSECV from 1-30 sampling runs, followed by a slight increase in the range of 30 to 400 runs of sampling. Finally, when the number of sampling runs exceeded 400, the RMSECV increased rapidly because some useful wavelengths were also eliminated. In Figure 6c, each curve records the coefficient path for each wavelength at different sampling runs, and the line of blue asterisks indicates the number of runs and the point of the lowest RMSECV value.

As a result of the CARS calculation, 38 wavelengths were selected from the full spectra for predicting black spot grades of the Yali pear. In comparison to the full spectra, the number of wavelengths decreased significantly. The selected wavelengths by CARS contained 429, 442, 444, 447, 460, 463, 465, 468, 476, 479, 489, 503, 505, 516, 518, 521, 537, 540, 558, 593, 596, 599, 626, 666, 696, 726, 797, 808, 911, 914, 916, 964, 967, 970, 972, 975, 989 and 992 nm, which were distributed throughout the original band range.

The spectra based on characteristic wavelengths by the CARS were the input variables used to establish the LS-SVM model; the performance of the model in terms of each grade is presented in the confusion matrix in Figure 5b. The predicted accuracies of the G0, G1, G2, and G3 samples were 97.30%, 94.59%, 94.59%, and

confusion matrix shown in Figure 5a. The predicted accuracies of the G0, G1, G2, and G3 samples were 97.30%, 94.59%, 97.30%, and 100%, respectively. The overall accuracy, precision, sensitivity, and specificity were 97.30%, 94.60%, 96.16%, and 98.21%, respectively. In comparison to the model based on the entire wavelengths, its predictive performance improved significantly. The results demonstrated an intrinsic relationship between the variables related to external perturbation selected using 2DCOS and the target grade classification. Additionally, model performance improvement indicates that the model can be effectively simplified, and its prediction ability can be improved by eliminating irrelevant variables.

97.30%, respectively. The overall accuracy, precision, sensitivity, and specificity were 95.95%, 88.35%, 94.60%, and 97.47%, respectively. The prediction performance is the same as that of the full-spectra model but is poorer than that of the 2DCOS.

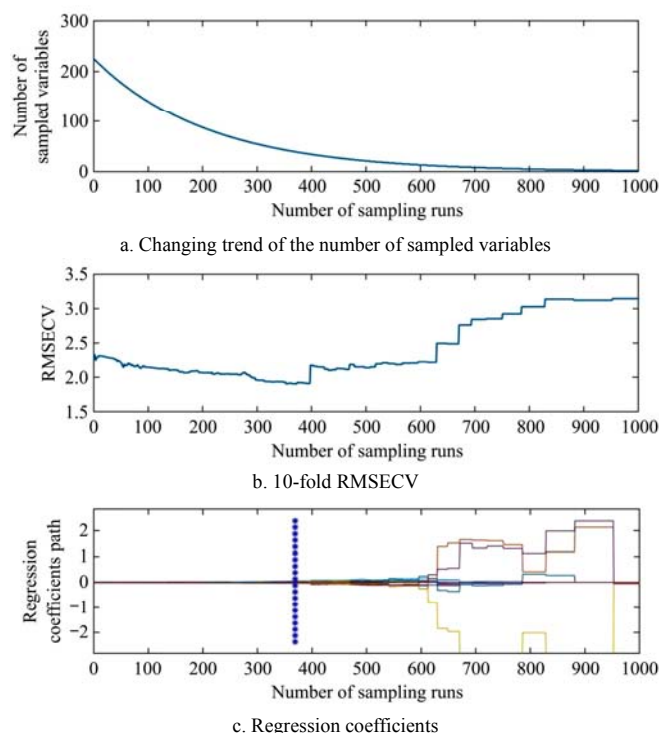


Figure 6 Wavelength selection by CARS

3.5.3 Model based on characteristic wavelengths using SPA

SPA was also used to select characteristic bands from the full bands. Multiple linear regression models of different subsets were established during the operation of the model, and the RMSEs of the models were calculated. The optimal wavelength can be determined from a subset of the minimum RMSE. Figure 7a shows the trend map of the change in the RMSE when variables of different wavelengths are selected. As shown in the data

presented in Figure 7a, the RMSE of the model reached a minimum of 1.3005 when 23 wavelength variables were selected. The number of wavelengths selected by the SPA was reduced to 23, compared to the former 224 bands, i.e., the number selected by the SPA amounts to only 10.3% of that in the full band. The distribution of the 23 optimal wavelengths selected by the SPA is shown in Figure 7b, which contained 402, 410, 418, 423, 434, 447, 458, 550, 556, 596, 634, 718, 797, 827, 872, 891, 914, 939, 970, 978, 986, 1000, and 1003 nm.

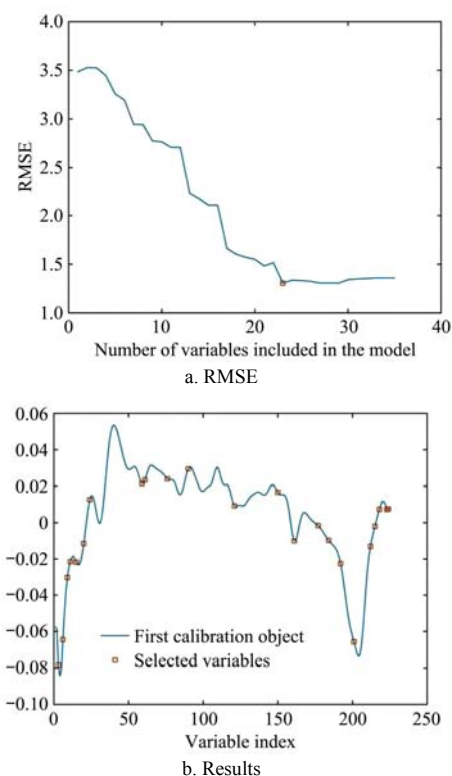


Figure 7 RMSE and results of SPA wavelength selection

After selecting the characteristic wavelengths by SPA, the LS-SVM model was established. The performance of the model in terms of each grade is presented in the confusion matrix in Figure 5c. The predicted accuracies of the G0, G1, G2, and G3 samples were 94.59%, 89.19%, 91.89%, and 97.30%, respectively. The overall accuracy, precision, sensitivity, and specificity were 93.24%, 86.31%, 87.46%, and 95.19%, respectively. The prediction performance was slightly lower than that of the full-spectra model. This may be due to the mistaken deletion of useful characteristic wavelengths.

4 Discussion

Pears have been threatened by black spot disease for a long time, which adversely affects their yield and quality. In this study, HSI technology was employed to detect and classify the different grades of corruption. The spectral data of black spot pears were used as the original information, and the effects of FD, SD, SNV and a combination of these on spectral noise removal were studied. A thorough comparison with previous individual pretreatment methods showed that an appropriate combination pretreatment method of mutually complementary methods could potentially improve model performance. However, the evidence provided in this study is not enough to prove the superiority of the combination method, and whether the combination method can optimize the spectrum of other fruits needs further verification.

The samples prepared at the same culture time also have

obvious differences due to clear individual differences in pears. Therefore, we used Sun's method for reference to conduct a secondary classification of the samples^[18], taking the lesion area as the classification standard rather than the incubation days. This can reduce the impact of individual differences.

In previous studies on fruit and vegetable detection, characteristic wavelength selection methods, such as CARS^[31], SPA^[32], and RF^[33], were employed, and the classification accuracy was significantly improved. However, these algorithms possessed low interpretability^[14], and the selected wavelengths may not be directly related to the physicochemical property changes. In this study, 2DCOS was attempted to select feature wavelengths, and the optimization effect was compared with CARS and SPA. The 2DCOS resulted in visible improvement in model performance, owing to its ability to discern small changes in the spectrum at the molecular level. This is consistent with the research conclusion of Yue et al.^[34]. In contrast with the commonly used algorithms, the 2DCOS method highlights the relationship between the characteristic wavelengths and chemical composition and explains the characteristic wavelengths from a chemical perspective, instead of a purely mathematical one.

5 Conclusions

In this study, the HSI technology was combined with 2DCOS to detect the black spot disease of Yali pears, and the decayed grades were accurately identified. The test results show that the LS-SVM model based on the spectra obtained by the SNV-SD-2DCOS algorithm yielded the best prediction performance. The identification accuracy of the prediction set was 97.30%, and the precision, sensitivity, and specificity were 94.60%, 96.16%, and 98.21%, respectively. In comparison to the model based on raw spectra and feature bands by the CARS and SPA, the 2DCOS model not only removed a large amount of redundant information but also selected a series of characteristic wavelengths from a chemical perspective. The results of the current study revealed that the 2DCOS algorithm is an effective tool for selecting feature-related wavebands of HSI data. This study lays a foundation for the development of HSI system for pathogenetic process monitoring and classification of fruit diseases. In the future work, more emphasis should be put on the accurate identification of the incubation period (infected, but no spot on the surface) of Yali black spot disease for early warning.

Acknowledgements

This work was financially supported by Hebei Province Key Research and Development Project (Grant No. 20327111D), Basic Scientific Research Funds of Hebei Provincial Universities (Grant No. KY202002), Key Laboratory of Modern Agricultural Engineering, Tarim University (Grant No. TDNG2020102), and the National Natural Science Foundation of China (Grant No. 31960498).

[References]

- [1] Han D H, Tu R L, Lu C, Liu X X, Wen Z H. Nondestructive detection of brown core in Chinese pear 'Yali' by transmission visible-NIR spectroscopy. *Food Control*, 2006; 17(8): 604–608.
- [2] Cao J K, Jiang W B. Induction of resistance in Yali pear (*Pyrus bretschneideri* Rehd.) fruit against postharvest diseases by acibenzolar-S-methyl sprays on trees during fruit growth. *Sci Hortic*, 2006; 110: 181–186.
- [3] Pan T T, Erkinbaev C, Sun D W, Jitendra P, Pu H B. Pathogenetic process monitoring and early detection of pear black spot disease caused by

- Alternaria alternata* using hyperspectral imaging. *Postharvest Biol Technol*, 2019; 154: 96–104.
- [4] Ali S, Mehdi K, Reza G M, Marziye M, Ahmad K H. Detection of foreign materials in cocoa beans by hyperspectral imaging technology. *Food Control*, 2021; 129: 108242. doi: 10.1016/j.foodcont.2021.108242.
- [5] Zhang B H, Li J B, Fan S X, Huang W Q, Zhao C J, Liu C L, Huang D F. Hyperspectral imaging combined with multivariate analysis and band math for detection of common defects on peaches (*Prunus persica*). *Computers and Electronics in Agriculture*, 2015; 114: 14–24.
- [6] Gómez-Sanchis J, Blasco J, Soria-Olivas E, Lorente D, Escandell-Montero P, Martínez-Martínez J M, et al. Hyperspectral LCTF-based system for classification of decay in mandarins caused by *Penicillium digitatum* and *Penicillium italicum* using the most relevant bands and non-linear classifiers. *Postharvest Biology and Technology* 2013; 82: 76–86.
- [7] Li J B, Luo W, Wang Z L, Fan S X. Early detection of decay on apples using hyperspectral reflectance imaging combining both principal component analysis and improved watershed segmentation method. *Postharvest Biology and Technology*, 2019; 149: 235–246.
- [8] Li J B, Huang W Q, Tian X, Wang C P, Fan S X, Zhao C. Fast detection and visualization of early decay in citrus using Vis-NIR hyperspectral imaging. *Comput Electron Agric*, 2016; 127: 582–592.
- [9] Zhang W W, Zhu Q B, Huang M, Guo Y, Qin J W. Detection and classification of potato defects using multispectral imaging system based on single shot method. *Food Analytical Methods*, 2019; 12: 2920–2929.
- [10] Wang W L, Li C Y, Tollner E W, Gitaitis R D, Rains G C. Shortwave infrared hyperspectral imaging for detecting sour skin (*Burkholderia cepacia*)-infected onions. *Journal of Food Engineering*, 2012; 109(1): 38–48.
- [11] Xuan G T, Gao C, Shao Y Y, Wang X Y, Wang Y X, Wang K L. Maturity determination at harvest and spatial assessment of moisture content in okra using Vis-NIR hyperspectral imaging. *Postharvest Biol Technol*, 2021; 180: 111597. Doi: 10.1016/j.postharvbio.2021.111597.
- [12] Yu H D, Qing L W, Yan D T, Xia G H, Zhang C H, Yun Y H, et al. Hyperspectral imaging in combination with data fusion for rapid evaluation of tilapia fillet freshness. *Food Chem*, 2021; 348: 129129. doi: 10.1016/j.foodchem.2021.129129.
- [13] Yun Y H, Li H D, Deng B C, Cao D S. An overview of variable selection methods in multivariate analysis of near-infrared spectra. *Trends Analyt Chem*, 2019; 113: 102–115.
- [14] Ding Y G, Zhang Q Z, Wang Y Z. A fast and effective way for authentication of *Dendrobium* species: 2DCOS combined with ResNet based on feature bands extracted by spectrum standard deviation. *Spectrochim Acta A Mol Biomol*, 2021; 261: 120070. doi: 10.1016/j.saa.2021.120070.
- [15] Wu H Y, Wei Y, Yang R J, Jin H, Ai C L. Analysis of chalk in rice by two-dimensional correlation spectroscopy. *J Mol Struct*, 2020; 1218: 128471. doi: 10.1016/j.molstruc.2020.128471.
- [16] Liu Y, Yao L Y, Xia Z Z, Gao Y G, Gong Z Y. Geographical discrimination and adulteration analysis for edible oils using two-dimensional correlation spectroscopy and convolutional neural networks (CNNs). *J Mol Struct*, 2021; 246: 118973. doi: 10.1016/j.saa.2020.118973.
- [17] Wang Y B, Zhang Y X, Zhang J H. Effects of salicylic acid on the behavior of Yali pear infected by *Alternaria kikuchiana* Tanaka. *Front Agric China*, 2010; 4: 79–83.
- [18] Sun Y, Xiao H, Tu S, Sun K, Pan L Q, Tu K. Detecting decayed peach using a rotating hyperspectral imaging testbed. *LWT*, 2018; 87: 326–332.
- [19] Huang H P, Hu X J, Tian J P, Jiang X N, Sun T, Luo H B, Huang D. Rapid and nondestructive prediction of amylose and amylopectin contents in sorghum based on hyperspectral imaging. *Food Chem* 2021; 359: 129954. doi: 10.1016/j.foodchem.2021.129954.
- [20] Gui J S, Fei J Y, Wu Z X, Fu X P, Diakite A. Grading method of soybean mosaic disease based on hyperspectral imaging technology. *Information Processing in Agriculture*, 2021; 8(3): 380–385.
- [21] Zhao N, Wu Z S, Cheng Y Q, Shi X Y, Qiao Y J. MDL and RMSEP assessment of spectral pretreatments by adding different noises in calibration/validation datasets. *Spectrochim Acta A Mol Biomol*, 2016; 163: 20–27.
- [22] Liang K, Huang J N, He R Y, Wang Q J, Chai Y Y, Shen M X. Comparison of Vis-NIR and SWIR hyperspectral imaging for the non-destructive detection of DON levels in *Fusarium* head blight wheat kernels and wheat flour. *Infrared Phys Technol*, 2020; 106: 103281. doi: 10.1016/j.infrared.2020.103281.
- [23] Ji Y M, Sun L J, Li Y S, Li J, Liu S C, Xie X, et al. Non-destructive classification of defective potatoes based on hyperspectral imaging and support vector machine. *Infrared Phys Technol*, 2019; 99: 71–79.
- [24] Li D R, Zhang F, Yu J S, Chen X F, Liu B J, Meng X H. A rapid and non-destructive detection of *Escherichia coli* on the surface of fresh-cut potato slices and application using hyperspectral imaging. *Postharvest Biol Technol*, 2021; 171: 111352. doi: 10.1016/j.postharvbio.2020.111352.
- [25] Mecozzi M, Pietroletti M, Mento R D. Application of FTIR spectroscopy in ecotoxicological studies supported by multivariate analysis and 2D correlation spectroscopy. *Vibrational Spectroscopy*, 2006; 44(2): 228–235.
- [26] Mokarizadeh H, Atashrouz S, Mirshekar H, Abdolhossein H S, Pour A M. Comparison of LSSVM model results with artificial neural network model for determination of the solubility of SO₂ in ionic liquids. *J Mol Liq*, 2020; 304: 112771. doi: 10.1016/j.molliq.2020.112771.
- [27] Akbar F M, Santosh L, Geonwoo K, Rahul J, Hoonsoo L, Sung K M, et al. Hyperspectral shortwave infrared image analysis for detection of adulterants in almond powder with one-class classification method. *Sensors*, 2020; 20(20): 5855. doi: 10.3390/s20205855.
- [28] Su W H, Yang C, Dong Y H, Johnson R, Page R, Szinyei T, et al. Hyperspectral imaging and improved feature variable selection for automated determination of deoxynivalenol in various genetic lines of barley kernels for resistance screening. *Food Chem*, 2021; 343: 128507. doi: 10.1016/j.foodchem.2020.128507.
- [29] McGlone V A, Kawano S. Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biol Technol*, 1998; 13(2): 131–141.
- [30] Gopal T, David C S Marita C. Nondestructive maturity determination in green tomatoes using a handheld visible and near infrared instrument. *Postharvest Biol Technol* 2013; 86: 221–229.
- [31] Jiang H Z, Jiang X S, Ru Y, Chen Q, Li X P, Xu L Y, Zhou H P, Shi M H. Rapid and non-destructive detection of natural mildew degree of postharvest *Camellia oleifera* fruit based on hyperspectral imaging. *Infrared Physics & Technology*, 2022; 123: 104169. doi: 10.1016/j.infrared.2022.104169.
- [32] Yuan R R, Liu G S, He J G, Wan G L, Fan N Y, Li Y, et al. Classification of Lingwu long jujube internal bruise over time based on visible near-infrared hyperspectral imaging combined with partial least squares-discriminant analysis. *Computers and Electronics in Agriculture*, 2021; 182: 106043. doi: 10.1016/j.compag.2021.106043.
- [33] Jiang Q Y, Wu G S, Tian C F, Li N, Yang H, Bai Y H, et al. Hyperspectral imaging for early identification of strawberry leaves diseases with machine learning and spectral fingerprint features. *Infrared Physics & Technology* 2021; 118: 103898. doi: 10.1016/j.infrared.2021.103898.
- [34] Yue J Q, Huang H Y, Wang Y Z. A practical method superior to traditional spectral identification: Two-dimensional correlation spectroscopy combined with deep learning to identify Paris species. *Microchemical Journal*, 2021; 160: 105731. doi: 10.1016/j.microc.2020.105731.