

Research on Detection Method of Pea Seed Vigor based on Hyperspectral Imaging Technology

Luo Qinjuan, Wang Lianming, Zhao Xiaoqing, Qian Hua, Yan Lei*

School of Technology, Beijing Forestry University,

Key Lab of State Forestry Administration for Forestry Equipment and Automation, Beijing 100083

China

Received: March 11, 2021. Revised: August 12, 2021. Accepted: August 24, 2021. Published: August 26, 2021.

Abstract— Rapid and noninvasive detection methods of seed vigor, an important index to evaluate seed quality, have been the research focus in recent years. In this paper, the detection method of pea seed vigor based on hyperspectral imaging technology was researched. First, the spectral images of different vigor grade samples with artificial aging were captured, and the original spectrum was pretreated with multiple scattering correction. Secondly, SPA and PCA were used to select respective bands. Finally, PLS-DA and LS-SVM model were established to identify the seed vigor of the pea seed, based on the whole band spectrum, the characteristic bands extracted by SPA and PCA respectively. The results showed that PLS-DA and LS-SVM models are effective, but LS-SVM had better performance. Through comparison, the method using full band spectrum was more accurate, the efficiency of method using 5 characteristic bands extracted by PCA was the highest while the way of extracting the representative band by SPA was the most meaningful to this study which achieved similar accuracy to the full band with only 20 bands. The SPA-LS-SVM method afforded the recognition accuracy (100%) for modeling set and validation set used to determine the vigor of pea seeds. The overall results suggest that hyperspectral imaging technology is useful for classification of different vitality pea seeds with non-

destructive manner, which can provide a basis for further development of online scoring devices.

Keywords—Hyperspectral imaging, seed vigor, characteristic band, pea, LS-SVM

I. INTRODUCTION

PEA is the second most important legume in the world and is recognized as a multi-purpose crop. The vigor of pea seeds directly determines whether pea production is efficient, stable, and reliable. In 1980, the North American Association of Official Seed Analysts defined seed vigor as: “Seed vigor refers to the potential for rapid seedling emergence and normal seedling growth under a wide range of field conditions” [1].

Many methods are used to measure seed vigor, which are roughly divided into two categories, such as traditional detection and new technology detection. Traditional methods include seedling growth measurements, germination rate measurements, accelerated aging test method, the TTC quantitative method, and conductivity measurements of the seed leaching solution [2]. The traditional detection method is relatively straightforward, and the indicators are clear, but there are inevitable shortcomings, such as damage to samples, time, and environmental pollution. The new technology detection method is simple and rapid. However, there are disadvantages,

such as the large measurement workload, poor repeatability, the long measurement period, the large environmental impact, and damage to seeds [3-10]. Due to these problems, fast, accurate, and non-destructive methods have become a research hotspot for seed workers. At present, to improve the healthy development of agriculture, improve the international competitiveness of Chinese agricultural products and alleviate environmental pressures, a simple, rapid, accurate, and non-destructive method for measuring seed vigor is needed. Hyperspectral imaging technology is about to become one of these new technologies. Compared with other physical measurement methods, hyperspectral technology produces a larger amount of information and is more complete, so it can comprehensively and better analyze seed vigor.

Hyperspectral imaging technology has used by some scholars to detect crop quality. Kailiang et al. explored the method of hyperspectral imaging technology applied to detect tea quality. The model identification rate reached 94%. Jiajia et al. used a hyperspectral imaging system and principal components analysis (PCA) after denoising and threshold segmentation using partial least squares regression to establish a model allowing for nondestructive testing of apples. Xing et al. used different stages of wheat hyperspectral images to observe the average emissivity and distinguish between normal wheat and budding wheat. Hyperspectral imaging technology has advantages of non-contact, high precision, high sensitivity, and fast real-time, easy to measure automation [11-14].

II. RESEARCH CONTENT AND TECHNICAL ROUTE

The technical route of this study is shown in Fig. 1. A rapid, non-destructive and accurate seed vigor detection method was explored using pea as the research object. First, a hyperspectral image acquisition system platform was built, and digital image processing technology combined with spectral analysis technology was used to detect the vitality characteristics of different peas. Second, a mathematical model of seed vigor level was established to explore the mechanism of rapid detection by hyperspectral imaging technology. Rapid, non-destructive testing of seed vigor was achieved.

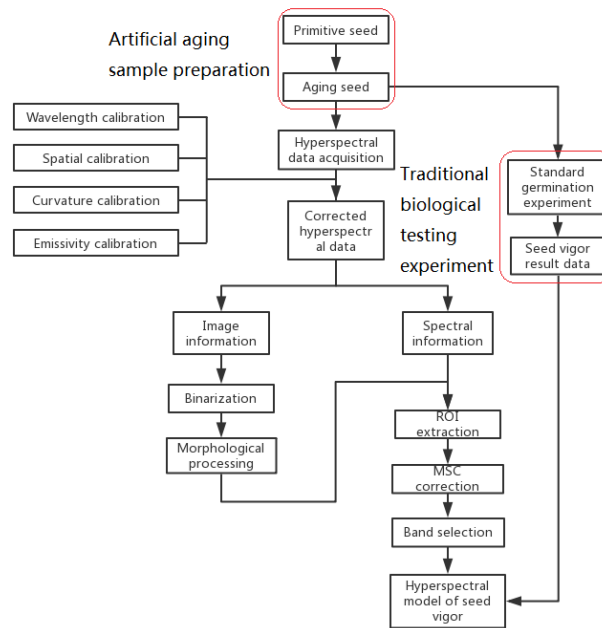


Fig. 1 Research technology route

A. Experimental and Data Collection

Hyperspectral imaging technology is an image data technology based on a very large number of narrow bands. It integrates advanced technologies in the fields of optics, electronics, information processing, and computer science. Hyperspectral imaging technology combines imaging technology with spectroscopic technology to detect two-dimensional targets. The geometric space and one-dimensional spectral information are combined to obtain continuous, narrow-band image data.

B. Hyperspectral Data Acquisition System

The hyperspectral data acquisition system used in the experiment is shown in Fig. 2. It consisted of a SOC710VP portable hyperspectral imager, a Schneider lens 17 mm/5.6, two 150 W halogen light sources, a sample, a Petri dish, a stage, and a PC. Two 150 W halogen lamps fixed on both sides of the stage served as the light source, providing stable continuous illumination in absolute darkness, and illuminating the surface of the sample (pea). A total of 128 bands of 400–1000 nm were collected with a hyperspectral imager. The pea seed image was then passed to the PC for storage and subsequent processing via a USB 2.0 data cable.

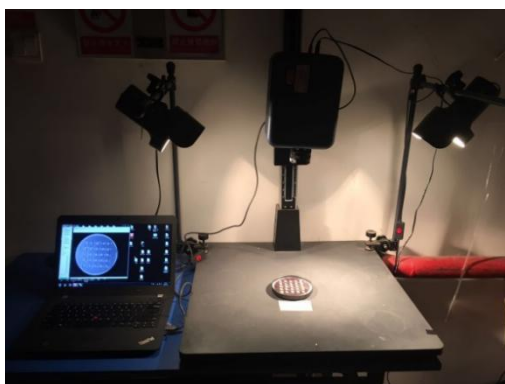


Fig. 2 Hyperspectral data acquisition system

C. Experimental Sample Preparation

In this study, pea seeds of different vitality levels were prepared by an artificial accelerated aging method to study sweet peas (growing period from September to December in Hainan, Guangdong, and other areas in spring and autumn). The peas were aged according to the pea seed technical standard (DB64/T 1045-2014).

Twenty-five randomly selected samples at room temperature ($20 \pm 0.5^\circ\text{C}$, humidity $60 \pm 0.5\%$) were considered category 1; 50 samples were selected for an artificial aging chamber and aged for 3 days at 60°C . From these, 25 samples were selected as category 2; the remaining 25 samples were aged for 3 days at 100°C and taken as category 3. The experimental sample is shown in Fig. 3.



Fig. 3 Experimental samples

D. Hyperspectral Data Acquisition

The three sample types subjected to different degrees of aging were placed in three Petri dishes. During the experiment, each dish was placed horizontally on a hyperspectral data acquisition platform, scanned by a hyperspectral analyzer at a rate of 23.2 sec/cube and 25 ms (after repeated adjustments) to obtain a three-dimensional (X, Y, λ) hyperspectral map. These images included a linear array scan of the detector in the Y direction and sample movement along the X direction. In this study, 520 pixels were obtained in the X direction, 696 pixels were obtained in the Y direction, and 128 bands were obtained

in the λ direction. During the experiment, the sample for image acquisition was maintained at room temperature. Hyperspectral photographs were taken every 30 min for 10 hours. The hyperspectral data of the three types of seeds were collected, abnormal data were deleted, and the sample data of intact pea hyperspectral $10 \times 3 \times 25$ or 1500 data were saved as a .cube file.

III. HYPERSPECTRAL IMAGE PREPROCESSING

A. Black and White Correction and Data Conversion

The intensity distribution of light is not uniform under different wavelength bands, and there is dark current in the sensor, resulting in a large amount of image noise in the band with weak light intensity, which requires image correction. Before the general spectral extraction step, two reference images (black reference and white reference) were used to correct the original hyperspectral image of pea seeds. The black reference image was obtained by turning off all lights and completely covering the lens with an opaque cover, while the white reference image was obtained by using white Teflon tiles with 100% high reflectivity. Finally, the corrected image can be calculated according to the following formula:

$$I = \frac{I_0 - B}{W - B} (3-1)$$

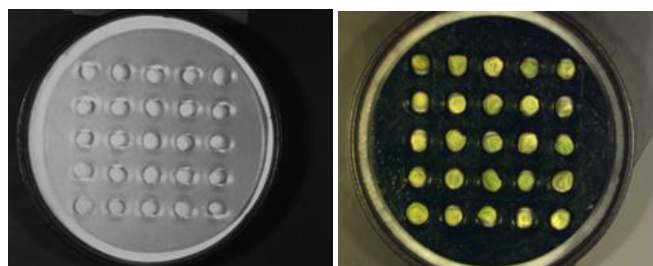
Where I ——Corrected hyperspectral image

I_0 ——Original hyperspectral image

W ——White reference image

B ——Black reference image

The image was corrected using the Spectral Radiance Analysis Toolkit, which is hyperspectral data processing software, and converted to a float file for processing by Matlab (MathWorks, Natick, MA, USA). The effect before and after calibration is shown in Fig. 4. The corrected data was used as the original hyperspectral data for subsequent processing.



a) Before correction

b) After correction

Fig. 4 RGB image before and after the black-white correction

B. Regional Segmentation

Some of the hyperspectral data were needed, and some were not. The methods used to obtain the data for this study are as follows:

- ① The binarization threshold was obtained using the optimal global threshold method;
- ② morphological processing was used to eliminate noise;
- ③ the connected area was obtained;
- ④ a binarized image was drawn after the open operation;
- ⑤ the center of gravity and region of interest (ROI) were drawn.

The original image was first binarized to extract the image feature values of the pea seeds. That is, the original gray values on the image were divided into two categories according to the size of the black (0) and white (1) threshold, so that the image could be changed to only black or white. The optimal global threshold method automatically calculates the segmentation threshold according to the distribution of the gray value in the image, improving adaptability of the algorithm [15].

The most basic method is to etch and swell the image for morphological transformation. This method effectively removes the outline of small impurities and facilitates the extraction of the contour in the pea seed image. In this study, the size and shape characteristic values of each seed in the image were extracted by morphological processing. Among them, corrosion and expansion are described by formulas (3-2) and (3-3):

Corrosion:

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad (3-2)$$

Swell:

$$A \oplus B = \{z | (\hat{B})_z \cap A \subseteq A\} \quad (3-3)$$

In the middle A —Image

B —Structural element

A and B are collections in Z^2

The opening and closing operations are based on corrosion and expansion, except that the order of corrosion and expansion differs. In the open operation, the operation process first corrodes, deletes some small noise points, and then expands to restore the original shape. Its effect is to smooth the image contour and eliminate small noise and spikes. In the closed

operation, the operation is first expanded and then etched, which is used to smooth the contour, fill the narrow discontinuities, and eliminate small holes in the image [16].

When acquiring a hyperspectral image of a sample, in addition to the required basic sample characteristics, it is often doped with unnecessary irrelevant information and noise, such as stray light, strong electrical noise, and spectral noise caused by operation and environmental influences. Since the interference information in the spectrum reduces the signal-to-noise ratio and the usefulness of the spectrum data, the addition of these noises will not be conducive to the analysis of the effective spectrum information of the sample. Generally, the extraction of effective ROI can reduce the influence of noise to a certain extent. In addition, the commonly used denoising method is spectral preprocessing [17 18]. In this study, multi-scatter corrections (MSC) were used to preprocess the spectra of pea seeds, which can effectively eliminate the spectral difference caused by different scattering levels, thereby enhancing the correlation between the spectrum and the data.

In the present study, the ROI region extracted a single pea seed, and the final result is shown in Fig. 5.

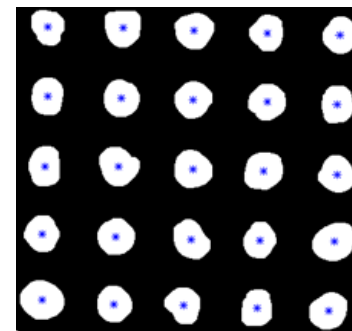


Fig. 5 Region of interest

C. Average spectrum

The average spectral value extracted from the ROI of each pea seed was used as the research object, and the average spectral curve of one of the samples obtained is shown in Fig. 6.

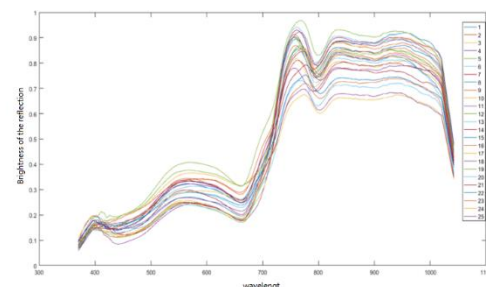


Fig. 6 Average spectrogram

As shown in Fig. 6, the spectral curves of the seeds in the same sample are basically the same; that is, the spectral curves

of different pea seeds of the same vitality had similar trends. The band noise before 400 nm and after 1,000 nm was relatively large. Therefore, the spectral curve in the range of 400–1000 nm was selected for study. In addition, the spectral curve between 450 and 570 nm rose relatively gently, whereas the spectral curve between 680 and 760 nm was steeper but tended to flatten after 810 nm.

D. Multiple Scattering Correction

MSC is the most widely used pretreatment technique in near-infrared spectroscopy, and it is still applicable in the field of hyperspectral imaging [19–20]. The steps to perform MSC preprocessing on the spectrum are as follows: (1) calculate the average spectrum of all sample spectra, and use this as the standard spectrum; (2) perform unary linear regression operation between the spectrum of each sample and the standard spectrum to obtain the linear translation and tilt offset of each spectrum relative to the standard spectrum; (3) subtract the linear translation amount from each original spectrum and divide it by the tilt offset to correct the relative tilt of the baseline of the spectrum. In this way, the baseline shift and offset of each spectrum are corrected by the reference standard spectrum.

The average spectrum of the calibration set was used as the reference spectrum in the present study. The corrected curve is shown in Fig. 7:

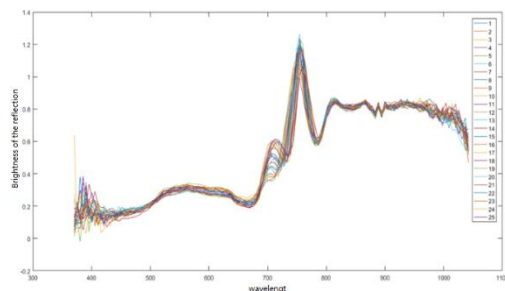


Fig. 7 Spectral curve after pretreatment by multi-scattering correction

The spectral intensity changed after the MSC correction. Although some obvious differences have been removed, the noise between 380–440 nm and 950–1040 nm has not been alleviated, so it is difficult to distinguish these data. The analysis will eliminate these two bands.

The comparison of the mean values of the normal, 60°C, 3-day, and 100-day 3-day samples after the multivariate scatter correction is shown in Fig. 8.

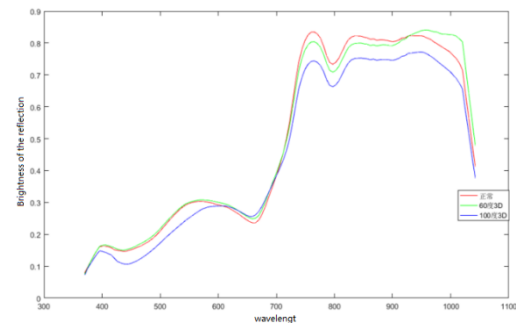


Fig. 8 Comparison of seed spectral curves of different vitality levels

As shown in Fig. 8, the spectral intensity changed after the MSC correction. Although some obvious differences have been removed, the noise between 380–440 nm and 950–1040 nm has not been alleviated, so there is still difficulty distinguishing these data. The analysis will eliminate these two bands.

The comparison of the mean values of the normal, 60°C, 3-day, and 100-day 3-day samples after multivariate scatter correction is shown in Fig. 8. Although the seeds of different vitality levels are basically consistent in hyperspectral direction, there are differences. According to the trend in the curve and the difference, the following conclusions can be drawn:

(1) The seed spectrum curves of different vitality levels were different, and seed vigor can be distinguished by spectral information, but the difference in the different vigor levels was not particularly obvious. Therefore, a qualitative model between the spectrum and the pea seeds of different vitality levels must be established if statistical methods are to be used, which will make it easier to analyze.

(2) The activity is better than normal seeds after aging at 60°C for 3 days before 700 nm and after 930 nm. The early stage may have occurred because this temperature promotes the activity or enzymes in the seeds. The medium-term trend reveals that the normal seed vigor level is higher. Due to the limitations of the hyperspectral band in the later stage, the mid-infrared and far-infrared bands cannot be seen, and the trend after predicting the spectra could not be obtained.

IV. EXPERIMENTAL DATA ANALYSIS

A. Successive Projection Algorithm Selects Feature Band

The collected hyperspectral information of the pea seeds is a three-dimensional data block composed of multiple bands, containing a large amount of redundant information and noise, with multi-collinearity and redundancy [21]. An effective means was needed to extract the effective wavelengths in the full band, to increase the modeling speed, and to reduce the size of the calculation. A variety of feature selection methods are used to extract the best wavelength that can represent all information

from the full-band model, and use this as the most effective wavelength. This can achieve the reduction of input features and dimensionality, which greatly speeds up the calculation speed, saves calibration time, and can effectively remove redundant information [22]. The Successive Projection Algorithm (SPA) is a forward selection algorithm and an effective band selection method. Previous studies have shown that SPA can effectively reduce collinearity between variables. The SPA consists of three main steps: selecting candidate variables based on the maximum projection values on the columns of the spectral matrix, evaluating all selected variables with root mean square error (RMSE), and deleting variables that are independent of the predicted properties.

The SPA can fully search for the set of variables in the spectral information with minimal redundant information and minimize the collinearity between the variables. At the same time, the number of variables used for modeling can be greatly reduced, and the speed and efficiency of modeling can be improved. The algorithm is briefly introduced as follows:

Let $X_{k(0)}$ be the initial iteration vector and N be the number of variables that need to be extracted. The spectral matrix is column j ;

a: Before starting the iteration, select one column j of the spectral matrix, assign the j th column of the modeling set to X_j , denoted as $X_{k(0)}$;

b: Record the set of unselected column vector positions as s , $s = \{j, 1 \leq j \leq J, j \notin \{X_{k(0)}, \dots, X_{k(n-1)}\}\}$

c: Calculate the projection of the remaining column vectors separately;

$$P_{x_j} = x_j - (x_j^T x_{k(n-1)}) x_{k(n-1)} (x_{k(n-1)}^T x_{k(n-1)})^{-1}, j \in s$$

d: Let $k_{(n)} = \tan^{-1}(\max(\|P_{x_j}\|)), j \in s$

e: Let $x_j = P_{x_j}, j \in s$

f: $n = n + 1$, if $n < N$, return to the b-cycle calculation.

Fig. 8 shows that the spectral curves of the seeds are the same in the range of the acquired spectral bands, but the reflection intensities of different seeds at the same wavelength differed, so the wavelength representing the seed vitality information cannot be determined. Therefore, it was necessary to determine the wavelength characteristics related to seed vigor.

The RMSE of different wavelengths selected after MSC preprocessing is shown in Fig. 9.

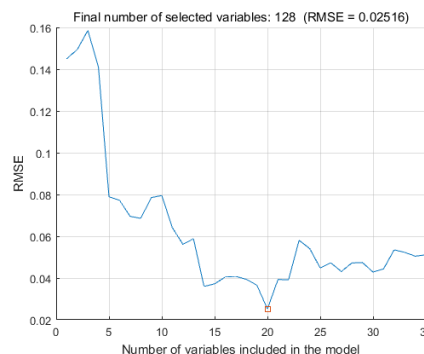


Fig. 9 Root means square error (RMSE) distribution diagram

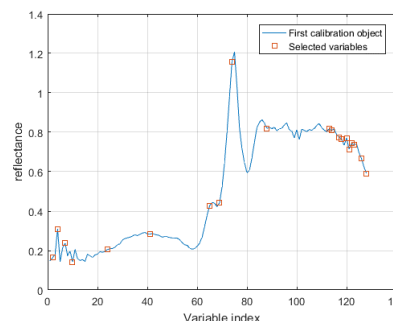


Fig. 10 Successive Projection Algorithm (SPA) selection wavelength

Twenty effective wavelengths were selected from Fig. 9, as shown in Fig. 10, where the curve represents the original spectrum and the box represents the selected effective wavelength. The selected bands were:

824, 749, 701, 722, 487, 1009, 998, 400, 987, 575, 1,015, 1,042, 415, 1031, 959, 981, 375, 385, 1,003, and 965 nm.

The unprocessed raw data were processed in the same way, and the comparative results are shown in Table 1.

Table 1 Comparison of effective band selection under the Successive Projection Algorithm (SPA)

	Number of bands	Effective band/nm	RMSE
RAW	8	643.318 959.752 659.092 685.459	0.21814
		701.323 436.215 749.120 722.529	
		824.075 749.120 1014.679	
		701.323 722.529 1042.283 487.428	
MSC	20	998.161 1009.169 981.677	0.02516
		415.835 1003.663 987.168	
		575.351 375.255	
		965.227 385.378 487.428 959.752 400.59	

As shown in Table 1, the RMSE (0.02516) value after MSC treatment is much smaller than the untreated RMSE (0.21814), so MSC can be used for pretreatment.

B. Principal Components Analysis

PCA is a dimensionality reduction tool that transforms the original variables with correlations in the direction of maximum covariance, and maps the high-dimensional spatial data transformed by coordinate transformation to a low-dimensional space to obtain the principal component, which is a new linearly uncorrelated variable. The original hyperspectral data were transformed by PCA and were regarded as a linear combination of the main component images. The proportion of the original image information of the principal component image was determined by the variance contribution rate. Generally, when the cumulative contribution rate of the principal component reaches a certain ratio, such as 85% or more, most of the hyperspectral data information can be explained. Therefore, the PCA-transformed hyperspectral data represents the original information to a large extent with only a small number of principal components, greatly reducing data processing time and eliminating redundant information in the original data. In this section, PCA was used to analyze the hyperspectral images with wavelengths ranging from 400 nm to 1,000 nm, and to find the main wavelengths that distinguish pea seed vigor.

The specific steps of the PCA were as follows:

(1) Standardize the raw data to eliminate the dimensional differences between variables:

The normalized matrix X is shown in Equation 4-1:

$$X = \begin{bmatrix} x_{11} & x_{21} & K & x_{1n} \\ x_{21} & x_{22} & L & x_{2n} \\ M & M & O & M \\ x_{n1} & x_{n2} & L & x_{nn} \end{bmatrix} \quad (4-1)$$

In the middle n—Number of sample points

The calculation for any X_{ij} in the matrix is as follows:

$$x_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (4-2)$$

In the middle \bar{X}_j —Average of the j th variable

s_j —Standard deviation of the j th variable

The formulas are shown as 4-3 and 4-4.

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad (4-3)$$

$$s_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}} \quad (4-4)$$

(2) The correlation coefficient matrix R of the normalized data matrix X was calculated:

$$R = \begin{bmatrix} r_{11} & r_{21} & K & r_{1p} \\ r_{21} & r_{22} & L & r_{2p} \\ M & M & O & M \\ r_{n1} & r_{n2} & L & r_{np} \end{bmatrix} \quad (4-5)$$

namely:

$$R = (r_{ij})_{p \times p} = X'X \quad (4-6)$$

In the middle p —The number of parameter variables in the sample; that is, the number of wavelengths collected

(3) The eigenvalue $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ of the correlation coefficient matrix R by $|R - \lambda I|$ and the corresponding unit eigenvector were determined:

$$u_1 = \begin{bmatrix} u_{11} \\ u_{21} \\ M \\ u_{p1} \end{bmatrix}, u_2 = \begin{bmatrix} u_{11} \\ u_{21} \\ M \\ u_{p2} \end{bmatrix}, u_p = \begin{bmatrix} u_{11} \\ u_{21} \\ M \\ u_{pp} \end{bmatrix} \quad (4-7)$$

(4) The main components are presented:

$$C_i = u_{i1}x_1 + u_{i2}x_2 + K + u_{ip}x_p \quad (4-8)$$

The variance contribution rate and the cumulative variance contribution rate corresponding to the principal component were calculated. The variance contribution rate α_i of C_i represented the percentage of the information contained in the principal component of the total information in the original data, and β_m represents the percentage of information retained by the first m principal component. This was calculated as follows:

$$\alpha_i = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k} \quad (4-9)$$

$$\beta_m = \frac{\sum_{i=1}^m \lambda_i}{\sum_{k=1}^p \lambda_k} \quad (4-10)$$

PCA processing was performed on the preprocessed hyperspectral data, and Fig. 11 shows the contribution rates of the first five principal components.

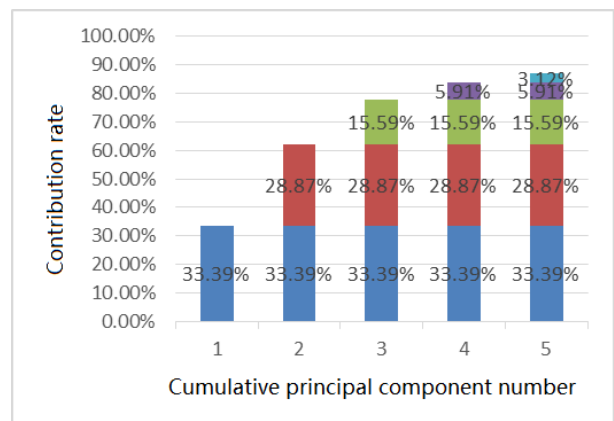


Fig. 11 Principal components analysis contribution coefficient for the full spectrum

Fig. 11 shows that the cumulative contribution rate of the first five principal components was 86.88%, which represents most of the hyperspectral data information, so the five principal components PC1–PC5 were selected.

Accordingly, PC1 was linearly superimposed by the spectral images of the 128 bands. To determine the main wavelengths affecting the expression of information in PC1, it was necessary to compare the weighted coefficients of the respective wavelengths to obtain the first principal component. The weighted coefficient distribution of each wavelength in the middle is shown in Fig. 12.

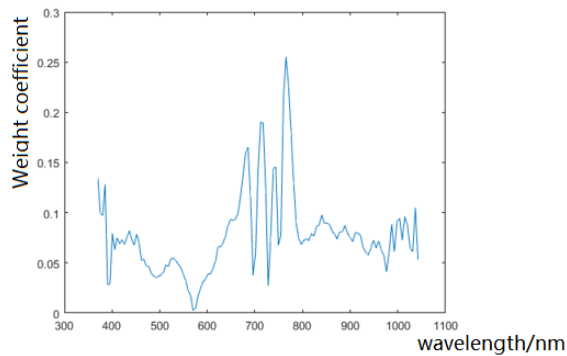


Fig. 12 Principal component 1 weight coefficient distribution

The PC1 weighted coefficients in the figure were processed as absolute values; that is, the weighted coefficients of the respective wavelengths were all positive values. Therefore, the wavelength at which the weighted coefficient is selected to be at the local maximum was the effective wavelength. Similarly, PC2, PC3, PC4, and PC5 were treated, and the effective wavelengths finally obtained were 370.2, 743.79, 754.45, 759.78, and 765.12 nm. This effective band has removed the same value of the local maximum of the weighted coefficients in each principal component.

V. ESTABLISHMENT AND EVALUATION OF THE SEED VIGOR CLASSIFICATION MODEL

A. Modeling Method for The Classification Model

Partial least squares discrimination analysis (PLS-DA) is a multivariate statistical analysis method of discriminant analysis. Discriminant analysis is a common statistical method to judge how to classify a research object based on the values of several observed or measured variables. The principle is to train the characteristics of different processing samples (such as observation and control samples) separately, generate training sets, and test the credibility of the training sets.

The PLS-DA explained that the number of sample observations is small, which reduced the influence of multicollinearity between variables. Multicollinearity refers to the strong linear relationship between some independent

variables. This situation is very common in practical applications, such as when studying the relationship between seed vigor and water, protein, starch, fat, and other factors. These independent variables are usually related. If the correlation is very high, it may be possible to use the least squares method to establish a regression equation.

The PLS-DA first treats the sample categories with response variables and is represented by the symbol δ , namely:

$$\delta_j = \begin{cases} 0 & \text{When } i \text{ samples belong to the } j\text{th class} \\ 1 & \text{When } i \text{ samples do not belong to the } j\text{th class} \end{cases} \quad (5-1)$$

Then, the partial least squares regression was used to establish a relational model between the explanatory variables and the response variables (dummy variables). Finally, the class of each sample was determined by comparing the predicted value of the response variable of the model; that is, if the predicted value of a certain dummy variable component is the largest, it is determined that the sample belongs to the category corresponding to the dummy variable.

Suykens proposed a new support vector machine method for solving classification and functional estimation problems called the least squares support vector machine (LS-SVM). The LS-SVM method uses the least squares linear system as the loss function instead of the quadratic programming method adopted by the traditional support vector machine. It simplifies computational complexity and is significantly faster than other SVM methods.

In this study, the LS-SVM method was used for classification. LS-SVM has similarities with SVM, and its essence is to find an optimal hyperplane to separate samples of different categories [23]. The difference between LS-SVM and SVM is that LS-SVM changes the inequality constraint of the original method into equality constraint, which greatly facilitates solution of the Lagrange multiplier. This calculation process effectively improves the calculation rate and the parameter optimization rate. But at the same time, LS-SVM loses the sparsity of the traditional SVM model. The original problem is QP (two problems of sub-planning and solving linear equations).

The original SVM problem is an inequality constraint with constraints:

$$\min_{\omega, b, \xi} J_p(\omega, \xi) = \frac{1}{2} \omega^T \omega + c \sum_{k=1}^N \xi_k \quad (5-2)$$

$$\text{such that } y_k [\omega^T \varphi(x_k) + b] \geq 1 - \xi_k, k = 1, \dots, N \\ \xi_k \geq 0, k = 1, \dots, N.$$

The LS-SVM problem is an equality constraint with the following constraints:

$$\min_{\omega, b, e} J_p(\omega, e) = \frac{1}{2} \omega^T \omega + \gamma \frac{1}{2} \sum_{k=1}^N e_k^2 \quad (5-3)$$

such that $y_k[\omega^T \varphi(x_k) + b] \geq 1 - e_k, k = 1, \dots, N$

The ξ in the original SVM problem is a slack variable, which is used to introduce outliers in the support vector. For the equality constraint of LS-SVM, the meaning of e and SVM on the right side of the equation is somewhat similar, and the final optimization target also includes e . In addition, in LS-SVM, the meaning of c in ξ SVM is the same. A weight is used to balance the search for the optimal hyperplane and the minimum amount of deviation. Next, similar to SVM, using the Lagrange multiplier method:

$$L(\omega, b, e, \alpha)$$

$$= J_p(\omega, e) - \sum_{k=1}^N \alpha_k \{y_k[\omega^T \varphi(x_k) + b] - 1 + e_k\} \quad (5-4)$$

The partial derivative was determined = 0 for α, b, e_k, α_k respectively:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{k=1}^N \alpha_k y_k \varphi(x_k) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^N \alpha_k y_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k, k = 1, \dots, N \\ \frac{\partial L}{\partial \alpha_k} = 0 \rightarrow y_k[\omega^T \varphi(x_k) + b] - 1 + e_k = 0, k = 1, \dots, N \end{cases} \quad (5-5)$$

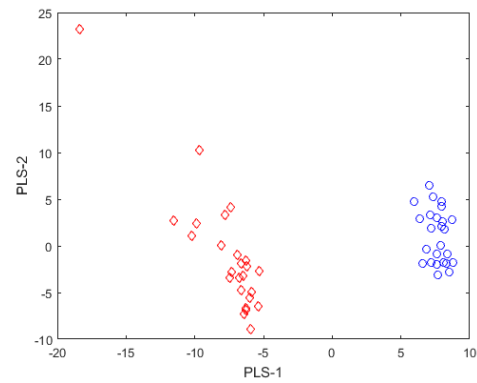
Next, according to these four conditions, a linear equation system for solving α and b is listed, and finally the LSSVM classification model is obtained:

$$y(x) = \text{sign}[\sum_{k=1}^N \alpha_k y_k K(x, x_k) + b] \quad (5-6)$$

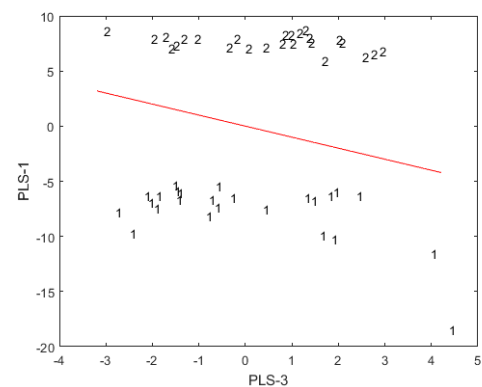
B. Seed vigor classification model based on PLS-DA

Based on the above content, this study uses the full band and the characteristic band extracted by PCA and SPA as input, and establishes a pea seed vigor discriminant model based on PLS-DA and LS-SVM. When modeling, the verification set and prediction set are randomly divided at a ratio of 3:1.

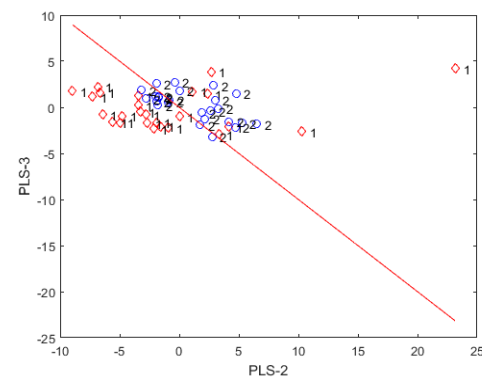
Fig. 13 shows the model established by PLS-DA in four different ways in the full spectral band. As shown in Fig. 13, the discrimination rates of PLS-1 and PLS-2 were 100%, and the discrimination rates of PLS-1 and PLS-3 were 100%, but PLS-2 and PLS-3 could not be completely distinguished, so the discrimination rate was only 58%, and the discrimination was determined by three variables; that is, the discrimination rates of PLS-1, PLS-2 and PLS-3 were still 100%.



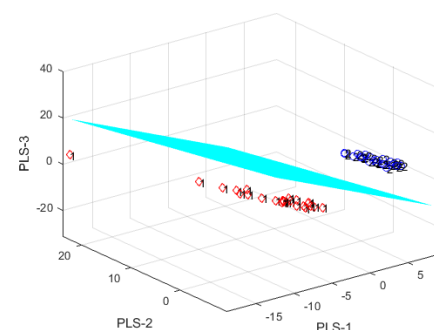
a) PLS-1 and PLS-2



b) PLS-1 and PLS-3



c) PLS-2 and PLS-3



d) PLS-1, PLS-2, and PLS-3

Fig. 13 The effect of model classification

Second, the 20 characteristic bands selected by PC1, PC2, PC3, PC4, PC5, and the SPA algorithm were used as input variables, and different prediction models were established by PLS-DA. The classification results are shown in Table 3.

Table 3 Discriminant results of the PLS-DA models

Methods	Variables	Calibration Set Accuracy	Prediction Set Accuracy
Raw-PLS-DA	128	100%	94.6%
PCA-PLS-DA	5	85.6%	65.5%
SPA-PLS-DA	20	99%	98%

Table 3 shows that the PLS-DA model has a higher classification accuracy rate for the pre-processed full-band spectral model and the PLS-DA model prediction set based on SPA extraction-based feature variables, and the latter is more accurate. The PLS-DA model based on the first five principal component feature variables obtained by the PCA analysis had the poorest prediction effect, and the classification accuracy of the prediction set was only 65.5%. As shown by the number of input variables, the more input variables involved in modeling, the better the prediction effect. When the pre-processed full-band spectral data was used and a total of 128 wavelength points were used as input variables, the correct rate of the prediction set was the highest when using full-band spectral modeling. The variables were large, the information volume was high, and the result was more accurate. When Feng et al. used SVM to establish a corn seed vigor recognition model, they also found that the recognition ability of the full-band model was better than that of the characteristic band model extracted by PCA [24]. However, the calculation was more complicated, less efficient, and contained more redundant information. Only a few variables were obtained by the PCA analysis, so it does not fully represent all the information of the original spectrum, but efficiency was the highest. The effective wavelength selected by the SPA algorithm fully represented the information of the original spectrum effectively, and the prediction effect was good. The SPA algorithm can be used as an effective means to extract wavelengths to improve the calculation speed of the model. The results show that it is feasible to use the SPA algorithm to select the effective feature wavelength and combine with the PLS-DA model to classify pea seed vigor grade and obtain satisfactory accuracy.

C. Seed vigor classification model based on LS-SVM

The prediction model was established using the LS-SVM model. The classification results are shown in Table 4. Compared with the PLS-DA model classification results, the LS-SVM model prediction results were significantly better than those of the PLS-DA model. The classification accuracy of the

model set and the prediction set of the pre-processed full-band spectrum, the first five principal component variables, and the 20 characteristic wavelength variables were all or nearly 100% using the LS-SVM model. The main reason is that the LS-SVM model is a nonlinear modeling method. In general, nonlinear modeling is more complicated compared with linear modeling, such as PLS-DA, and nonlinear influencing factors are considered during the model building process; thus, improving the prediction of the model. The results show that the LS-SVM classification effect is better, and the effective feature band post-modeling after SPA extraction simplified the model, which fully represented the information of the original spectrum effectively, and was an effective means of extracting the wavelength to improve the model operation speed.

In summary, PLS-DA and LS-SVM can be used to distinguish three different vigorous pea seeds. When using the hyperspectral visible/near-infrared reflectance imaging system to classify infected and healthy watermelon seeds, the recognition accuracy of PLS-DA and LS-SVM can reach more than 90% [25]. Although the full-band performance of the two models was better than that of the characteristic band model, when the SPA characteristic band was used as input, similar effects can be achieved with fewer variables. This fully demonstrates the good recognition ability when hyperspectral imaging technology is combined with the discriminant model.

Table 4 Discriminant results of LS-SVM models

Methods	Variables	Calibration Set Accuracy	Prediction Set Accuracy
Raw-LS-SVM	128	100%	100%
PCA-LS-SVM	5	98.6%	98%
SPA-LS-SVM	20	100%	100%

VI. CONCLUSION

Detection of pea seed vigor is a very complicated process. Many external and internal factors affect pea seed vigor. This study has verified the superiority of hyperspectral detection technology, and obtained the following conclusions:

(1) Noise was reduced, random variations were eliminated, and the processed data were close to the ideal spectral data by using MSC multi-scattering pre-processed data.

(2) The SPA selected 20 effective bands, and the PCA selected five effective bands. A large number of hyperspectral data bands was obtained, but higher real-time performance was realized with a smaller number of bands used for online detection. Extracting a finite number of characteristic wavelengths reduced the data dimensions, reduced the amount of input, improved efficiency, and provided a theoretical basis for designing a rapid detection device for the handheld rice germination rate based on the limited characteristic wavelength.

(3) The PLS-DA and LS-SVM prediction models were established based on the full-band spectral variables, the first five principal component variables obtained by the PCA analysis, and the characteristic wavelength variables selected by the SPA algorithm. The discriminant results of the PLS-DA and LS-SVM models based on characteristic wavelengths were better than those based on the principal component variables. The LS-SVM model based on feature wavelengths had the best recognition effect. The 20 characteristic wavelength variables selected by SPA better reflected the spectral information compared with the PCA method, and the classification accuracy of the modeling set and the prediction set were all close to 100%.

The test results show that applying hyperspectral technology quickly and accurately identified pea seed vigor.

In the future, we will study many different pea seeds and establish a more stable and applicable pea seed viability discrimination model to further improve the reliability of determining pea seed vigor.

VII. DISCUSSION

In this study, hyperspectral imaging technology was used to distinguish and identify the three aging grades of pea seeds. It can be seen from the average spectrum that the spectra of seeds of different aging grades have similar changing trends, and there were peaks and valleys near the same band. And as the aging time increased, the overall reflectivity showed a downward trend. These results fully proved that hyperspectral imaging technology can achieve qualitative analysis of different grades of pea seeds.

Before applying PLS-DA and LS-SVM to achieve quantitative discrimination, PCA and SPA were used to select characteristic bands. The experimental results show that the non-linear model LS-SVM has higher ability to distinguish pea seeds than the linear model PLS-DA, which may be due to the large overlap of the spectral information of the three aging grade seeds. Secondly, because all the information is included in the full-band model, the accuracy of the model is the highest. The feature band extracted by SPA corresponds to the location of the peak and valley, so it can achieve better discrimination compared to the PCA feature model.

Based on the above results, when applying hyperspectral imaging technology to achieve non-destructive testing, the nonlinear model is more suitable for multi-category samples that are difficult to distinguish. Appropriate preprocessing methods and feature selection methods can reduce the amount of calculation and improve the performance of the model.

ACKNOWLEDGMENT

This study was financially supported by National Natural Science Foundation of China (Grant No. 31770769) and the

Fundamental Research Funds for the Central Universities (NO.2015ZCQ-GX-03).

References

- [1] International Rules for Seed Testing. International Seed Testing Association. Seed Science and Technology . 1985
- [2] Jia PANG, Hong LIU, Haiqing ZHANG, Aimin LIU, Jiwei He, Ye LIU, Junlong LIU, Yixin ZHANG. Effect of Corona Treatment and Dielectric Separation on Seed Vigor of Hybrid Rice[J]. Agricultural Biotechnology, 2018, 7(04):48-54.
- [3] Sweet Pepper Seed Inspection Using Image Processing Techniques[J]. Prarin Chupawa, Kiattisin Kanjanawanishkul. Advanced Materials Research . 2014 (931)
- [4] A novel auto-sorting system for Chinese cabbage seeds. Huang K, Cheng J. Sensors . 2017
- [5] Identification of paddy varieties based on novel seed angle features. Chaugule A A, Mali S N. Computers and Electronics in Agriculture . 2016
- [6] R.A. Braga Jr, I.M. Dal Fabbro, F.M. Boremetal. Assessment of seed viability by laser speckle techniques[J]. Biosystems Engineering, 2003, 86(3):287~294.
- [7] Artur Zdunek, Anna Adamiak, Piotr M. The biospeckle method of agricultural crop: a review. Optics and Laser Engineering, 52(2014):276~285.
- [8] Fernández M, Mavilio A, Rabal H, et al. Characterization of Viability of Seeds by Using Dynamic Speckles and Difference Histograms[M]// Progress in Pattern Recognition, Speech and Image Analysis. Springer Berlin Heidelberg, 2003:329-333.
- [9] Baranowski P, Mazurek W, Walczak R T. Use of Thermography for Pre-sowing Evaluation of Seed Germination Capacity[J]. Acta Horticulturae, 2003, 6004(2):459-465.
- [10] Kranner I, Kastberger G, Hartbauer M, et al. Noninvasive diagnosis of seed viability using infrared thermography.[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(8):3912.
- [11] Gaikwad S, Chitre A V, Dandawate Y H. Classification of Indian Classical Instruments Using Spectral and Principal Component Analysis Based Cepstrum Features[P]. , 2014.
- [12] Deleforge, Antoine, Forbes, Florence, Horaud, Radu. Hyper-spectral image analysis with partially-latent regression[P]. Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European, 2014.

- [13] Sicong Liu, Bruzzzone, L., Bovolo, F., Peijun Du. A novel sequential spectral change vector analysis for representing and detecting multiple changes in hyperspectral images[P]. Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, 2014.
- [14] Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks[J] . Gamal ElMasry, Ning Wang, Clément Vigneault. Postharvest Biology and Technology . 2008 (1)
- [15] Krishna, M.P., Sriram, A., Puhan, N.B.. Clustering based image binarization in palm leaf manuscripts[P]. Advance Computing Conference (IACC), 2014 IEEE International, 2014.
- [16] Guo Yanyan, Guo Qi. A Class of Gray Morphological Expansion and Corrosion Operators[J]. Chinese Quarterly Journal of Mathematics, English, 2017, 32(02):181-186.
- [17] Baranowski P, Mazurek W, Pastuszka-Woźniak J. Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data[J]. Postharvest Biology & Technology, 2013, 86(2013):249-258.
- [18] Kamruzzaman M, Elmasry G, Sun D W, et al. Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging.[J]. Food Chemistry, 2013, 141(1):389-96.
- [19] H. Martens, S.A. Jensen, P. Geladi, Multivariate linearity transformations for near infrared reflectance spectroscopy, in: O.H.J. Christie (Editor), Proc. Nordic Symp. Applied Statistics, Stokkland Forlag, Stavanger, Norway, 1983, pp. 205–234.
- [20] P. Geladi, D. MacDougal, H. Martens, Appl. Spectrosc. 39 (1985) 491.
- [21] V. Ponomaryov, A. Palacios-Enriquez. Sparse Approach in Filtering of Color Images Corrupted by Mixture Noises[J]. WSEAS Transactions on Signal Processing, 2020, 16:81-88.
- [22] S. Muni Rathnam, G. Siva Koteswara Rao. A Novel Deep Learning Architecture for Image Hiding[J]. WSEAS Transactions on Signal Processing, 2020, 16:206-210.
- [23] S. Komeylian. Optimization Modeling of the Hybrid Antenna Array for the DoA Estimation[J]. International Journal of Electronics and Communication Engineering, 2021, 3: 75-80.
- [24] L. Feng, S. Zhu, C. Zhang, Y. Bao, X. Feng, Y. He. Identification of Maize Kernel Vigor under Different Accelerated Aging Times Using Hyperspectral Imaging[J]. Molecules, 2018, 23: 3078.
- [25] H. Lee, M. S. Kim, Y. Song, C. Oh, H. Lim, Wang. Lee, Jum. Kange, B. Cho. Non-destructive evaluation of bacteria-infected watermelon seeds using visible/near-infrared hyperspectral imaging[J]. Journal of the Science of Food and Agriculture, 2017, 97: 1084-1092.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US