

Research Article

Sensitive Wavelengths Selection in Identification of *Ophiopogon japonicus* Based on Near-Infrared Hyperspectral Imaging Technology

Zhengyan Xia,¹ Chu Zhang,¹ Haiyong Weng,¹ Pengcheng Nie,^{1,2} and Yong He¹

¹College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

²State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China

Correspondence should be addressed to Pengcheng Nie; npc2012@zju.edu.cn and Yong He; 101006238@qq.com

Received 21 March 2017; Accepted 29 May 2017; Published 27 August 2017

Academic Editor: Shaoping Li

Copyright © 2017 Zhengyan Xia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hyperspectral imaging (HSI) technology has increasingly been applied as an analytical tool in fields of agricultural, food, and Traditional Chinese Medicine over the past few years. The HSI spectrum of a sample is typically achieved by a spectroradiometer at hundreds of wavelengths. In recent years, considerable effort has been made towards identifying wavelengths (variables) that contribute useful information. Wavelengths selection is a critical step in data analysis for Raman, NIRS, or HSI spectroscopy. In this study, the performances of 10 different wavelength selection methods for the discrimination of *Ophiopogon japonicus* of different origin were compared. The wavelength selection algorithms tested include successive projections algorithm (SPA), loading weights (LW), regression coefficients (RC), uninformative variable elimination (UVE), UVE-SPA, competitive adaptive reweighted sampling (CARS), interval partial least squares regression (iPLS), backward iPLS (BiPLS), forward iPLS (FiPLS), and genetic algorithms (GA-PLS). One linear technique (partial least squares-discriminant analysis) was established for the evaluation of identification. And a nonlinear calibration model, support vector machine (SVM), was also provided for comparison. The results indicate that wavelengths selection methods are tools to identify more concise and effective spectral data and play important roles in the multivariate analysis, which can be used for subsequent modeling analysis.

1. Introduction

Ophiopogon japonicus (the tuber of *Ophiopogon japonicus* Ker-Gawl., Liliaceae) was originally documented in “Shennong materia medica,” in which it was classified as high grade [1]. The herb is sweet, slightly bitter, and slightly cold, enters the heart, lung, and stomach channels, moisturizes lung by nourishing yin, purges heat, promotes the production of body fluids, and so forth. It is one of the most commonly used Chinese medicines, which is widely applied to clinic. *Ophiopogon japonicus* is mainly planted in Zhejiang and Sichuan province in China, recognized as “Zhemaiddong” and “Chuanmaiddong” in China, respectively [2]. And there exist great differences in growing and harvesting between them as follows: “Zhemaiddong” is three-year cultivated, while “Chuanmaiddong” is one-year cultivated. On soil conditions, “Zhemaiddong” is planted coastally, while “Chuanmaiddong”

is planted inland. Therefore, although “Zhemaiddong” and “Chuanmaiddong” are similar in appearance, they not only differ in quality, but also have a large difference in price. The traditional identification methods of *Ophiopogon japonicus* of different growing areas include thin layer chromatography, high performance liquid chromatography, gas chromatography, and liquid chromatography-mass spectrometry. These analytical methods are generally complex and time-consuming and require consumption of chemical reagents and higher professional requirements for operators; therefore, it is necessary to develop a rapid and nondestructive identification method of “Zhemaiddong” and “Chuanmaiddong.”

Hyperspectral imaging (HSI) technology has emerged as an alternative technique that can meet both spatial and spectral requirements and thus has been widely applied in quality evaluation and classification of Traditional Chinese

Medicine. Zhang et al. fabricated a visible-near-infrared (Vis-NIR) HSI portable field spectrometer to distinguish sun-dried and sulfur-fumigated Chinese medicine herbs and achieved the results with a sensitivity of 96.4% and a specificity of 98.3% for RPA identification [3]. Tankeu et al. classified *Stephania tetrandra* and the nephrotoxic *Aristolochia fangchi* based on hyperspectral imaging. A discrimination model with a coefficient of determination (R^2) of 0.9 and a root mean square error of prediction (RMSEP) of 0.23 was created [4]. The potential of short wave infrared (SWIR) hyperspectral imaging and image analysis as a rapid quality control method to distinguish between *Illicium anisatum* (Japanese star anise) and *Illicium verum* whole dried fruit was investigated. A classification model with 4 principal components and an $R^2 X_{cum}$ of 0.84 and $R^2 Y_{cum}$ of 0.81 was developed for the 2 species using partial least squares-discriminant analysis (PLS-DA). The model was subsequently used to accurately predict the identity of *I. anisatum* (98.42%) and *I. verum* (97.85%) introduced into the model as an external dataset [5]. Sandasi et al. applied HIS, MIR, and NIR spectroscopy to certify ginseng reference materials and commercial products. And good discrimination models with high $R^2 X$ and Q^2 cum values were developed [6]. These results suggest that hyperspectral imaging is a potential technique to control medicine quality for medical applications.

A HSI spectrum of a sample is typically measured by a spectroradiometer for hundreds of wavelengths. The large number of spectral variables in most spectral datasets often renders the prediction of a dependent variable unreliable. However, the use of appropriate projection or selection techniques, such as principle component analysis or partial least squares regression, may minimize this problem [7]. Recently, considerable efforts have been made on developing and evaluating different programs that identify variables that contribute useful information or eliminate variables that contain redundancy data. The basic principle of the selection method is to select a small number of representative variables from the original set of variables. And the purpose of variable selection is to select a subset of spectral variables that produce the smallest possible errors when used to establish determination or classification models. Variable selection is an important step in multivariate analysis because the removal of redundant variables will produce better prediction results [8].

Two to five wavelengths selection methods were usually compared [9–11], however, which could not fully reflect the effectiveness and importance of wavelength selection methods in multivariate analysis. In the present work, 10 wavelength selection methods were compared in classification of “Zhemaidong” and “Chuanmaidong” to discuss the application of wavelength selection methods in multivariate analysis.

2. Materials and Methods

2.1. Materials. A total of 675 *Ophiopogon japonicus* samples were collected, of which 315 samples were purchased from

different growers of Cixi, Zhejiang province, and 360 samples were derived from different growers of Mianyang, Sichuan province.

2.2. Hyperspectral Imaging System. A hyperspectral imaging system was used in the experiment, which consists of an imaging spectrograph (Impector V10E, Spectral Imaging Ltd., Oulu, Finland), a CCD camera (C8484-05, Hamamatsu city, Japan), a lens (OLE-23, Specim, Spectral Imaging Ltd., Oulu, Finland), an illuminant source with two quartz tungsten halogen lamps (Fiber-Lite DC950, Dolan Jenner Industries Inc., Boxborough, USA), a conveyer belt controlled by a stepper motor (IRCP0076 Isuzu Optics Corp, Taiwan, China), and a computer. The whole system was assembled in a dark chamber except the computer, as shown in Figure 1.

2.3. Acquisition and Calibration of Hyperspectral Images. After repeated tests, the height between the lens and the sample was set as 15 cm, the exposure time of camera was set as 1.35 ms, and the speed of the conveyer was set as $18.7 \text{ mm}\cdot\text{s}^{-1}$. The hyperspectral image was acquired by a software (Spectral Image-V10E, Isuzu Optics Corp, Taiwan, China).

The acquired raw hyperspectral images should be calibrated with the white and dark reference according the following equation:

$$R = \frac{I_{\text{raw}} - I_{\text{black}}}{I_{\text{white}} - I_{\text{black}}}, \quad (1)$$

where R was the corrected image, I_{raw} was the raw hyperspectral image, I_{white} was the white reference with nearly 100% reflectance acquired by the special white Teflon tile, and I_{black} was the dark reference image with nearly 0% reflectance obtained by turning off the light source together with covering the camera lens.

2.4. Wavelengths Selection Methods

2.4.1. Successive Projections Algorithm, SPA. Successive projections algorithm (SPA) is an efficient method of spectral feature selection, which could minimize the collinearity between variables [12].

2.4.2. Regression Coefficient Method, RC. Regression coefficient is calculated based on PLS, and sensitive wavelengths are usually selected according to the regression coefficient of the optimal PLS model. Generally, the peaks or bands where the absolute value of RC is greater than threshold are selected as sensitive wavelength or waveband [9].

2.4.3. Loading Weights Method, LW. The loading weights show the importance of corresponding wavelength or bands in the spectral matrix. The peaks or valleys with the maximum absolute loading weights from the first principal factor to the optimal principal factor are selected as sensitive wavelengths [13, 14].

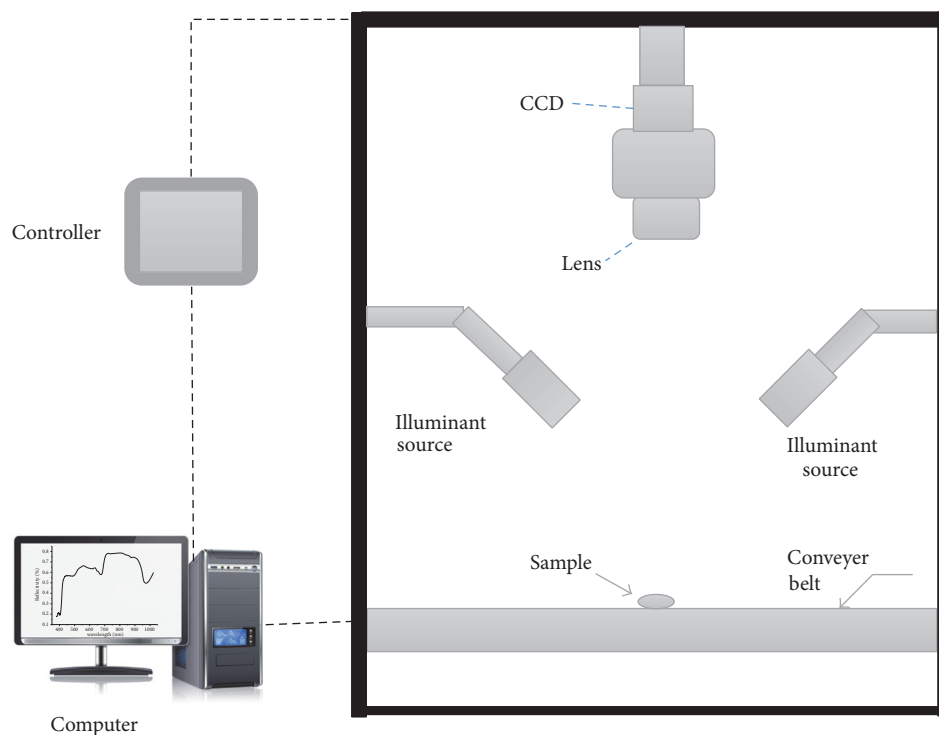


FIGURE 1: The hyperspectral imaging system.

2.4.4. Uninformative Variable Elimination, UVE. Uninformative variable elimination (UVE) is widely applied for variable selection based on analysis of the regression coefficients of the PLS model. It can eliminate noninformative variables and the remaining is useful for the chemical and classification analysis [15, 16].

2.4.5. Competitive Adaptive Reweighted Sampling, CARS. Competitive adaptive reweighted sampling (CARS) is a feature variable selection method combining Monte Carlo sampling with PLS regression coefficient. Adaptive reweighted sampling is employed in CARS, and the variables with larger weight of regression coefficient are applied as a new subset to establish PLS model, and after repeated calculation, the subset with the lowest root mean square error of cross validation (RMSECV) is chosen [17, 18].

2.4.6. Interval PLS, iPLS. In the iPLS method, the data are divided into nonoverlapping sections; each section develops a separate PLS model to identify the most useful variable range [19, 20].

2.4.7. Backward Interval PLS, BiPLS. For the backward iPLS (BiPLS) algorithm, the dataset is split into a given number of intervals; the PLS models are then calculated with each interval left out in a sequence; that is, if n intervals are chosen, then each model is based on $n - 1$ intervals that exclude one interval at a time. The first omitted interval gives the poorest performing model with respect to RMSECV [21, 22].

2.4.8. Forward Interval PLS, FiPLS. As in the interval PLS model, the dataset is split into a given number of intervals, but the PLS models are then developed based on successively improving intervals with respect to RMSECV; that is, if n intervals are used, then the first model is based on one interval that has the best performance, the second model uses the next interval, and so on [23, 24].

2.4.9. Genetic Algorithm-PLS, GA-PLS. The method combines the advantage of GA and PLS and is the most commonly used method for spectral data analysis. GA applied to PLS have been shown to be very efficient optimization procedures. They have been applied on many spectral datasets and have been proved to provide better results than full-spectrum methods [25, 26].

2.4.10. UVE-SPA Method. In this method, UVE eliminates uninformative variables, and then SPA is employed for variable selection. Fewer variables are selected by a UVE-SPA algorithm compared to UVE.

2.4.11. Model Evaluation and Software. The efficiency of the wavelengths selection method is based on the identification rate and the number of variables. The efficiency equation is as follows:

$$E = \frac{(D_s - D_f) \times (N_f - N_s)}{N_f} \times 100, \quad (2)$$

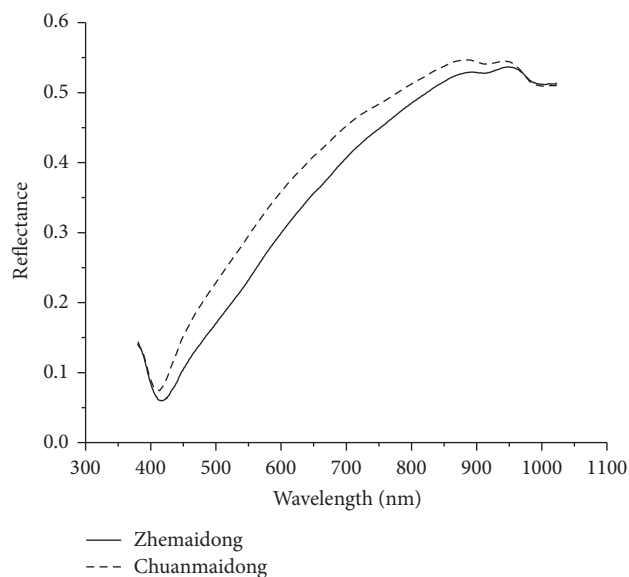


FIGURE 2: Average raw spectra reflectance curves of *Ophiopogon japonicus*.

where E was the efficiency of wavelengths selection method, D_s was the identification rate of prediction set in the model established by variables selected by the wavelengths selection method, and D_f was the identification rate of prediction set in the full-spectrum model. N_f was the number of variables of full-spectrum and N_s was the number of variables selected by wavelength selection method.

When $E > 0.5$, the wavelength extraction method is proved to be highly efficient.

When $-0.5 \leq E \leq 0.5$, the method is proved to be efficient, except when $E = 0$, $(N_f - N_s)/N_f \geq 0.8$; the method is proved to be highly efficient.

When $E < -0.5$, the method is proved to be of low efficiency.

The spectral data extraction, SPA, UVE, UVE-SPA, iPLS, BiPLS, FiPLS, CARS, GA-PLS, and SVM were conducted on Matlab R 2010b (The Math Works, Natick, MA, USA). LW, RC, and PLS-DA were performed on Unscrambler® 10.1 (CAMO AS, Oslo, Norway).

3. Results and Discussion

3.1. Raw Spectra Reflectance Curves of *Ophiopogon japonicus*.

The spectra of “Zhemaidong” and “Chuanmaidong” were acquired in the range of 380–1030 nm. The raw average spectra of “Zhemaidong” and “Chuanmaidong” were shown in Figure 2. No significant differences were observed in the range of 380–401 nm and 961–1030 nm, while different magnitudes of the spectra reflectance could be found in the range of 402–960 nm. Wavelength selection methods were further employed to identify feature information for better classification of “Zhemaidong” and “Chuanmaidong.”

675 *Ophiopogon japonicus* samples were split into two sets, calibration set and prediction set. “Zhemaidong” samples were labeled as “1,” while “Chuanmaidong” samples were

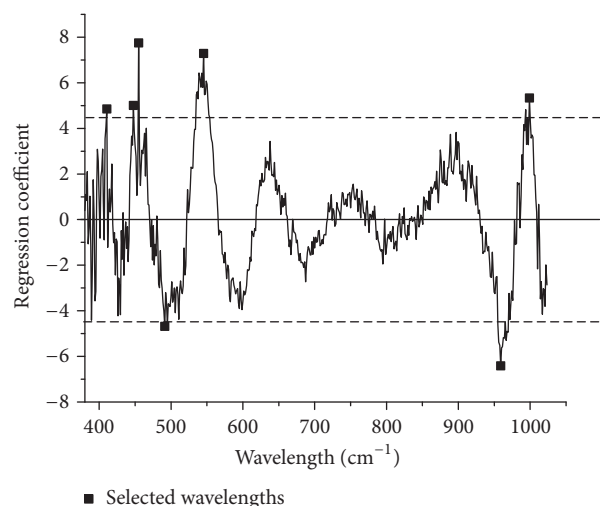


FIGURE 3: Seven wavelengths were selected by RC method.

TABLE 1: Class assignment and division of *Ophiopogon japonicus* samples.

	Zhemaidong	Chuanmaidong
Label	1	2
Calibration set	210	240
Prediction set	105	120
Sum up	315	360

labeled as “2.” *Ophiopogon japonicus* samples were divided in Table 1.

3.2. Sensitive Wavelengths Selection. Firstly, each wavelengths selection method should be optimized to evaluate the performance of each method better.

3.2.1. SPA. The number of sensitive wavelengths was set as 5~30, and 5 wavelengths (889, 1014, 411, 460, and 407 nm) were selected.

3.2.2. RC. Regression coefficients of PLS model based on full-spectrum were shown in Figure 3. The threshold was set as ± 4.5 ; finally, 7 wavelengths were selected by RC method.

3.2.3. UVE. UVE method was applied for full-spectrum data with no pretreatment; the number of principal components was set as 20. The selection criteria of threshold were 99% of the maximum value of variable stability. 291 wavelengths were selected by UVE (as shown in Figure 4), in which columns represent selected wavelengths.

3.2.4. UVE-SPA. 291 variables were selected by UVE method; SPA was applied to minimize the number of variables selected by UVE. 12 wavelengths were extracted finally which were shown in Figure 5.

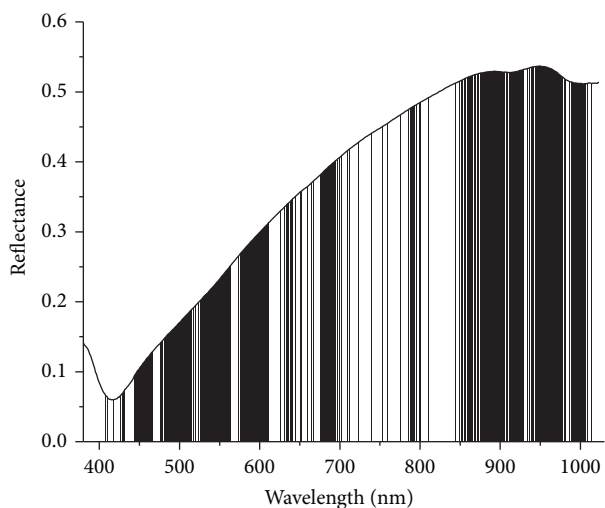


FIGURE 4: Plot of 291 selected wavelengths by UVE. Columns represent selected wavelengths.

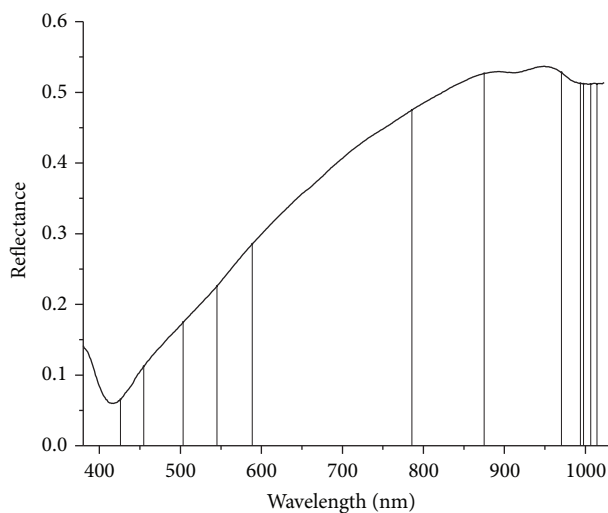


FIGURE 5: Plot of 12 selected wavelengths by UVE-SPA. Columns represent selected wavelengths.

3.2.5. *iPLS*. The raw dataset was split to 16–32 intervals, and the optimal interval was selected according to the lowest root-mean-squares error of cross validation (RMSECV). As shown in Figure 6, 16 intervals were the optimum mode. RMSECV of PLS models based on 16 intervals, respectively, were shown in Figure 7; the 15th interval was selected as the sensitive wavelengths range.

3.2.6. *BiPLS*. The optimal result was achieved when raw dataset was divided into 32 intervals. Finally, seen in Figure 8, 13 intervals were selected as follows, numbers 3, 5, 7, 9–11, 13, 14, 24, 25, and 30–32.

3.2.7. *GA-PLS*. In the GA-PLS method, population size was set as 30, probability of mutation was set as 0.01, probability of

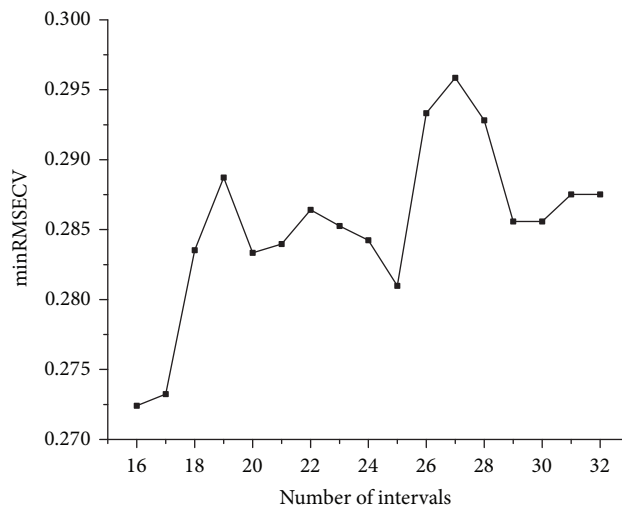


FIGURE 6: Minimum RMSECV of different number of intervals.

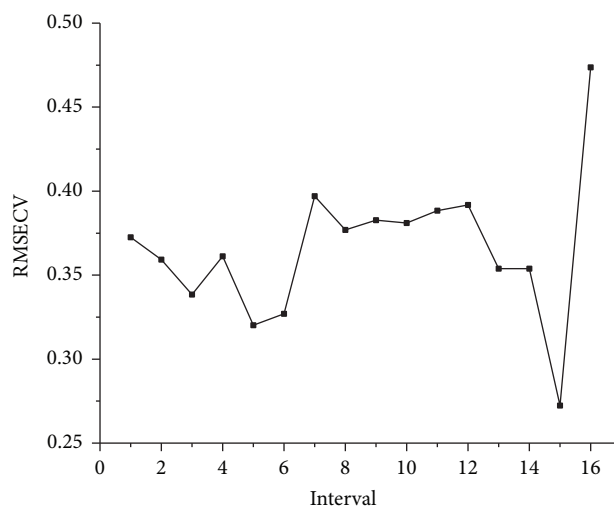


FIGURE 7: RMSECV of 16 intervals.

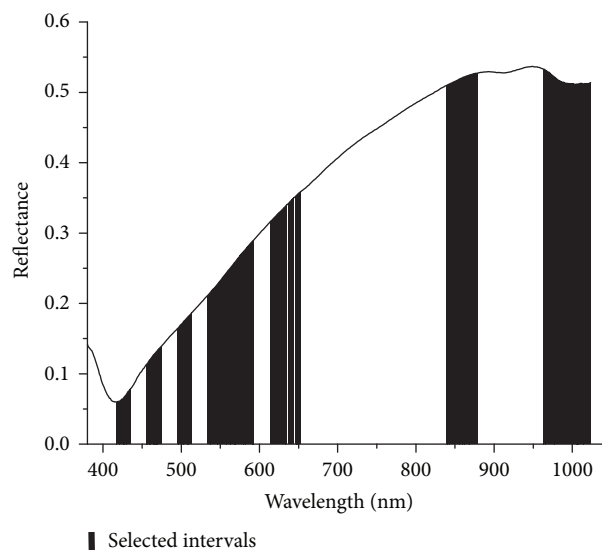


FIGURE 8: 13 intervals selected by BiPLS. Columns represent selected intervals.

cross-over was set as 0.5, and the number of runs was chosen to be 100. 85 wavelengths were selected eventually.

After optimization procedure, wavelengths selected by different methods were shown in Table 2. The number of wavelengths selected by SPA, RC, LW, and UVE-SPA method was under 12, while that by iPLS and GA-PLS method was 32 and 85, respectively, and that by UVE, CARS, and BiPLS method was no more than half of the full-spectrum, whereas FiPLS eliminated only 32 redundant variables.

3.3. Partial Least Squares-Discriminant Analysis (PLS-DA) Model. PLS-DA models were established based on variables selected by different methods and the raw full-spectrum, respectively, and prediction results of different models were compared (shown in Table 3). In UVE-PLS-DA, UVE-SPA-PLS-DA, CARS-PLS-DA, BiPLS-PLS-DA, FiPLS-PLS-DA, and GA-PLS-PLS-DA models, the identification accuracy improved and the number of variables reduced. Seen from Table 3, the identification accuracy of all models was over 88%. Compared with PLS-DA model based on the raw dataset, the identification accuracy of BiPLS-PLS-DA model increased from 95.1% to 99.1%, while the number of variables decreased from 512 to 208, which stated BiPLS was an effective wavelengths selection method. Although the identification accuracy of PLS-DA models based on SPA, RC, LW, and iPLS method was lower than that of full-spectrum, the number of variables greatly decreased. Compared with PLS-DA model based on full-spectrum, the number of variables of SPA-PLS-DA model decreased 99%, from 512 to 5, while the identification accuracy only decreased 1.8%. The number of variables of RC-PLS-DA model was reduced 98.6%, while the identification accuracy only decreased 1.3%. The number of variables of PLS models based on LW and iPLS models decreased 98.4%, while the identification accuracy decreased only 6.7%. According to the efficiency of each wavelengths selection method, SPA, RC, LW, and iPLS were methods of low efficiency, FiPLS was efficient method, and UVE, UVE-SPA, CARS, BiPLS, and GA-PLS were highly efficient methods.

3.4. Support Vector Machine (SVM) Models. Support vector machine is suitable for solving small sample, nonlinear, and high dimensional pattern problems. SVM models were developed based on variables selected by different methods and full-spectrum, and the results were compared. Seen from Table 4, the identification accuracy of all models was over 88%. The optimal performance was achieved by GA-PLS-SVM model; the identification accuracy of calibration set and prediction set were 99.6% and 99.1%, respectively. Compared with SVM model established with raw spectra, the performance of UVE-SVM, BiPLS-SVM, and FiPLS-SVM models was better, and the number of variables reduced with different degrees. Although the identification accuracy of UVE-SPA-SVM and CARS-SVM decreased 0.5%, the number of their variables reduced 97.7% and 79.5%, respectively. Similarly, the identification accuracy of SPA-SVM, RC-SVM, LW-SVM, and iPLS-SVM models was worse than that of raw spectra, with a decrease of no more than 8.5%, while

the number of variables reduced 99.0%~93.8%. Therefore, wavelengths selection which greatly improves the operation rate and prediction effect is an extraordinary step of modeling analysis. According to the efficiency of each model, SPA, RC, LW, and iPLS were methods of low efficiency, UVE, UVE-SPA, CARS, and FiPLS were efficient method, and BiPLS and GA-PLS were highly efficient methods.

3.5. Comparison of Different Models. According to the discriminant results of SVM and PLS-DA models, BiPLS and GA-PLS were highly efficient methods. SPA, RC, LW, and iPLS were methods of low efficiency, probably because these methods greatly reduce variables but also remove some useful information. The efficiency of UVE, UVE-SPA, and CARS was different in the two models. Above all, in identification of “Zhemaiddong” and “Chuanmaiddong,” BiPLS and GA-PLS were efficient wavelengths selection methods.

Seen from Tables 3 and 4, the identification accuracy of all models was above 88%. The average identification accuracy of prediction set of PLS-DA models and SVM models based on different selected variables was 94.8% and 95.1%, respectively, which stated the similar performance of PLS-DA models and SVM models in identification of *Ophiopogon japonicus*. The optimal performance of PLS-DA and SVM models was achieved based on BiPLS and GA-PLS methods, respectively, with the identification accuracy over 99%. The results indicated that using variable selection methods to select sensitive wavelengths was efficient for reduction of spectral data as well as the establishment of classification model, and it was feasible to identify “Zhemaiddong” and “Chuanmaiddong” by applying near-infrared hyperspectral imaging technology.

4. Conclusions

In this study, in view of hyperspectral data of *Ophiopogon japonicus*, 10 wavelengths selection methods were chosen to explore the general characteristics of different feature selection methods. UVE, UVE-SPA, CARS, BiPLS, and GA-PLS were highly efficient methods when the selected variables were used to develop PLS-DA models, and CARS achieved the best performance where the number of variables reduced 79.5% and the identification accuracy increased from 95.1% to 98.2%. Of all the SVM models based on different selected variables, BiPLS and GA-PLS were highly efficient methods, and GA-PLS performed the best where the number of variables decreased 83.4% and the identification accuracy increased from 96.9% to 99.1%, which was consistent with the literatures [26, 27]. SPA, RC, LW, and iPLS were methods of low efficiency; the number of variables decreased greatly while the identification accuracy reduced slightly, while Zhang et al. proved that, in the determination of soluble protein content in oilseed rape leaves based on near-infrared hyperspectral imaging, of all the sensitive wavelengths selection method, SPA performed better than GA-PLS and RC [28]. Therefore, the performance of each method could be different according to different spectral data.

TABLE 2: Effective wavelengths selected by different methods.

Methods	Number	Wavelengths/nm
SPA	5	889, 1014, 411, 460, 407
RC	7	409, 448, 455, 491, 545, 959, 999
LW	8	550, 990, 433, 1014, 539, 385, 380, 382
UVE	291	408, 410, 417, 425, 428, 430, 432, 444~467, 477~479, 481~517, 519, 524~564, 574, 576~611, 626, 631~639, 642, 644, 650, 652, 659, 664, 667, 676~695, 697, 700, 703, 709, 711, 723, 740, 752, 759, 776, 786, 788~796, 799, 800, 810, 844, 849, 852, 853, 856, 857, 859~929, 934, 937, 940~981, 986, 990~1007, 1009, 1014
UVE-SPA	12	426, 455, 503, 545, 589, 786, 875, 970, 994, 998, 1007, 1014
CARS	105	418, 426, 431, 437, 443, 457, 466, 475~481, 489, 491~495, 497~500, 507, 511, 516, 519, 533, 535, 539, 540, 543~545, 548~551, 555, 558, 565~569, 571, 573, 576, 582, 584, 594~610, 613~618, 624, 637, 640, 643, 648, 653, 661, 668, 675, 681, 685, 687, 689, 719, 735, 738, 750, 751, 795, 806, 813, 816, 831, 856, 862, 874~877, 881, 888~889, 905, 910, 918, 920, 924, 961~964, 968, 973, 976, 987, 992~996, 1023
iPLS	32	942~982
BiPLS	208	418~436, 456~474, 494~513, 534~592, 614~653, 839~879, 963~1023
FiPLS	480	418~1023
GA-PLS	85	431, 466~469, 472, 473, 479~481, 490~494, 506~509, 512~522, 534~550, 551, 580~584, 685~689, 799~801, 875~877, 888, 956~963, 965~978, 983, 985, 990~1000

TABLE 3: Results of PLS-DA models using different selected wavelengths.

Methods	Variables	Calibration		Prediction		<i>E</i>
		Correct number	Identification accuracy/%	Correct number	Identification accuracy/%	
Raw	512	449	99.8	214	95.1	
SPA	5	417	92.7	210	93.3	-1.78
RC	7	413	91.8	211	93.8	-1.28
LW	8	402	89.3	199	88.4	-6.60
UVE	291	449	99.8	219	97.3	0.95
UVE-SPA	12	449	99.8	216	96.0	0.88
CARS	105	449	99.8	221	98.2	2.46
iPLS	32	430	95.6	199	88.4	-6.28
BiPLS	208	450	100	223	99.1	2.38
FiPLS	480	449	99.8	219	97.3	0.14
GA-PLS	85	446	99.1	217	96.4	1.08

TABLE 4: Results of SVM models using different wavelengths.

Methods	Variables	Calibration		Prediction		<i>E</i>
		Correct number	Identification accuracy/%	Correct number	Identification accuracy/%	
Raw	512	449	99.8	218	96.9	-4.46
SPA	5	431	95.8	208	92.4	-3.55
RC	7	429	95.3	210	93.3	-5.71
LW	8	415	92.2	205	91.1	0.17
UVE	291	448	99.6	219	97.3	-0.49
UVE-SPA	12	448	99.6	217	96.4	-0.40
CARS	105	444	98.7	217	96.4	-7.97
iPLS	32	442	98.2	199	88.4	0.77
BiPLS	208	444	98.7	221	98.2	0
FiPLS	480	444	98.7	218	96.9	1.83
GA-PLS	85	448	99.6	223	99.1	

The study indicated that the wavelengths selection method could extract a small number of variables containing effective information and eliminating noninformation variables. Variables selection methods were tools to identify more concise and effective spectral data and played important roles in the multivariate analysis, which could be used for subsequent modeling analysis. Meanwhile, the characteristic wavelengths selected could provide a theoretical basis for the development of instruments.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant nos. 61405175 and MOI201702), the National Key Research and Development Program of China (Grant no. 2016YFD0700304), and the National Key Scientific Instrument and Equipment Development Project (Grant no. 2014YQ47037702).

References

- [1] National pharmacopoeia committee, "Pharmacopoeia of People's Republic of China," p. 155, 2015, National pharmacopoeia committee, Pharmacopoeia of People's Republic of China.
- [2] J. Zhang and Y. Wang, "Determination of Total Flavonoids and Total Saponins in *Ophiopogon japonicus* (Thunb.) Ker. Gawl by NIR Spectroscopy," *Chinese Journal of Spectroscopy Laboratory*, vol. 29, no. 01, pp. 551–555, 2012.
- [3] H. Zhang, T. Wu, L. Zhang, and P. Zhang, "Development of a Portable Field Imaging Spectrometer: Application for the Identification of Sun-Dried and Sulfur-Fumigated Chinese Herbs," *Applied Spectroscopy*, vol. 70, no. 5, pp. 879–887, 2016.
- [4] S. Tankeu, I. Vermaak, W. Chen, M. Sandasi, and A. Viljoen, "Differentiation between two "fang ji" herbal medicines, *Stephania tetrandra* and the nephrotoxic *Aristolochia fangchi*, using hyperspectral imaging," *Phytochemistry*, vol. 122, pp. 213–222, 2016.
- [5] I. Vermaak, A. Viljoen, and S. W. Lindström, "Hyperspectral imaging in the quality control of herbal medicines - The case of neurotoxic Japanese star anise," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 75, pp. 207–213, 2013.
- [6] M. Sandasi, I. Vermaak, W. Chen, and A. Viljoen, "The application of vibrational spectroscopy techniques in the qualitative assessment of material traded as ginseng," *Molecules*, vol. 21, no. 4, article no. 472, 2016.
- [7] Z. Xiaobo, Z. Jiewen, M. J. W. Povey, M. Holmes, and M. Hanpin, "Variables selection methods in near-infrared spectroscopy," *Analytica Chimica Acta*, vol. 667, no. 1–2, pp. 14–32, 2010.
- [8] R. M. Balabin and S. V. Smirnov, "Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data," *Analytica Chimica Acta*, vol. 692, no. 1–2, pp. 63–72, 2011.
- [9] H. Z. Chen and Z. Y. Xia, "Determination of total flavonoids in propolis based on NIR spectroscopy technology," *Chinese Journal of Pharmaceutical Analysis*, vol. 34, no. 10, pp. 1868–1873, 2014.
- [10] P. Nie, D. Wu, D.-W. Sun, F. Cao, Y. Bao, and Y. He, "Potential of visible and near infrared spectroscopy and pattern recognition for rapid quantification of notoginseng powder with adulterants," *Sensors (Basel, Switzerland)*, vol. 13, no. 10, pp. 13820–13834, 2013.
- [11] P. Nie, Z. Xia, D.-W. Sun, and Y. He, "Application of visible and near infrared spectroscopy for rapid analysis of chrysin and galangin in Chinese propolis," *Sensors (Basel, Switzerland)*, vol. 13, no. 8, pp. 10539–10549, 2013.
- [12] C. Zhang, F. Liu, W. Kong, H. Zhang, and Y. He, "Fast identification of watermelon seed variety using near infrared hyperspectral imaging technology," *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, vol. 29, no. 20, pp. 270–277, 2013.
- [13] Y. Wang, Y. Gao, X. Yu, Y. Wang, S. Deng, and J.-M. Gao, "Rapid Determination of *Lycium Barbarum* Polysaccharide with Effective Wavelength Selection Using Near-Infrared Diffuse Reflectance Spectroscopy," *Food Analytical Methods*, 2015.
- [14] Y. Shao, C. Xie, L. Jiang, J. Shi, J. Zhu, and Y. He, "Discrimination of tomatoes bred by spaceflight mutagenesis using visible/near infrared spectroscopy and chemometrics," *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, vol. 140, pp. 431–436, 2015.
- [15] L. Pan, R. Lu, Q. Zhu, K. Tu, and H. Cen, "Predict Compositions and Mechanical Properties of Sugar Beet Using Hyperspectral Scattering," *Food and Bioprocess Technology*, vol. 9, no. 7, pp. 1177–1186, 2016.
- [16] J. Dong, W. Guo, Z. Wang, D. Liu, and F. Zhao, "Nondestructive Determination of Soluble Solids Content of 'Fuji' Apples Produced in Different Areas and Bagged with Different Materials During Ripening," *Food Analytical Methods*, vol. 9, no. 5, pp. 1087–1095, 2016.
- [17] S. Fan, B. Zhang, J. Li, W. Huang, and C. Wang, "Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple," *Biosystems Engineering*, vol. 143, pp. 9–19, 2016.
- [18] B.-C. Deng, Y.-H. Yun, D.-S. Cao et al., "A bootstrapping soft shrinkage approach for variable selection in chemical modeling," *Analytica Chimica Acta*, vol. 908, pp. 63–74, 2016.
- [19] Y. Zhao, S. Wang, Z. Li, F. Cao, and Z. Pei, "A Novel Interval Integer Genetic Algorithm Used for Simultaneously Selecting Wavelengths and Pre-processing Methods," *Chinese Journal of Analytical Chemistry*, vol. 44, no. 9, pp. e1609–e1616, 2016.
- [20] X. H. Chen and J. Huang, "Wavelength selection algorithm based on iPLS and CARS multi-model fusion technology," *Computer Engineering and Application*, vol. 52, no. 16, pp. 229–232, 2016.
- [21] X. Zou, J. Zhao, and Y. Li, "Selection of the efficient wavelength regions in FT-NIR spectroscopy for determination of SSC of 'Fuji' apple based on BiPLS and FiPLS models," *Vibrational Spectroscopy*, vol. 44, no. 2, pp. 220–227, 2007.
- [22] B. Mahanty, S.-U. Yoon, and C.-G. Kim, "Spectroscopic quantitation of tetrazolium formazan in nano-toxicity assay with interval-based partial least squares regression and genetic algorithm," *Chemometrics and Intelligent Laboratory Systems*, vol. 154, pp. 16–22, 2016.
- [23] Q. Yang, G. H. Zhu, P. Ren, S. Long, and J. D. Yang, "Wavelength Selection for NIR Spectroscopic Analysis of Chemical Oxygen Demand Based on Different Partial Least Squares Models," *Journal of Analytical Science*, vol. 32, no. 4, pp. 485–489, 2016.

- [24] F. Qu, D. Ren, J. Hou et al., The Characteristic Spectral Selection Method Based on Forward and Backward Interval Partial Least Squares, Spectroscopy and Spectral Analysis,.
- [25] X.-B. Ding, C. Zhang, F. Liu, X.-L. Song, W.-W. Kong, and Y. He, "Determination of soluble solid content in strawberry using hyperspectral imaging combined with feature extraction methods," *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis*, vol. 35, no. 4, pp. 1020–1024, 2015.
- [26] C. Zhang, N. Xu, L. Luo et al., "Detection of Aspartic Acid in Fermented Cordyceps Powder Using Near Infrared Spectroscopy Based on Variable Selection Algorithms and Multivariate Calibration Methods," *Food and Bioprocess Technology*, vol. 7, no. 2, pp. 598–604, 2014.
- [27] L. Jiao, S. Bing, X. Zhang, Y. Wang, and H. Li, "Determination of enantiomeric composition of tryptophan by using fluorescence spectroscopy combined with backward interval partial least squares," *Analytical Methods*, vol. 7, no. 11, pp. 4535–4540, 2015.
- [28] C. Zhang, F. Liu, W. Kong, and Y. He, "Application of visible and near-infrared hyperspectral imaging to determine soluble protein content in oilseed rape leaves," *Sensors (Switzerland)*, vol. 15, no. 7, pp. 16576–16588, 2015.