American Journal of Analytical Chemistry, 2022, 13, 63-77
https://www.scirp.org/journal/ajac
ISSN Online: 2156-8278
ISSN Print: 2156-8251

Geographical Traceability of *Clinacanthus nutans* with Near-Infrared Pectroscopy and Chemometrics

Fengyan Yu[1], Jinfang Ma[2], Yi Qi[3,4], Han Song[2], Guiliang Tan[5], Furong Huang[2*], Maoxun Yang[1,4*]

[1]Guangdong Provincial Key Laboratory of Research and Development of Natural Drugs, the Second Clinical Medical College, School of Pharmacy, Guangdong Medical University, Dongguan, China
[2]Opto-Electronic Department of Jinan University, Guangzhou, China
[3]The Marine Biomedical Research Institute of Guangdong Zhanjiang, Zhanjiang, China
[4]Marine Chinese Medicine Branch, National Engineering Research Center for Modernization of Traditional Chinese Medicine, Zhanjiang, China
[5]School of Material Science and Food Engineering, University of Electronic Science and Technology of China, Zhongshan Institute, Zhongshan, China
Email: *furong_huang@jnu.edu.cn, *yangmaoxun@gdmu.edu.cn

How to cite this paper: Yu, F.Y., Ma, J.F., Qi, Y., Song, H., Tan, G.L., Huang, F.R. and Yang, M.X. (2022) Geographical Traceability of *Clinacanthus nutans* with Near-Infrared Pectroscopy and Chemometrics. *American Journal of Analytical Chemistry*, **13**, 63-77.
https://doi.org/10.4236/ajac.2022.132006

Received: January 7, 2022
Accepted: February 25, 2022
Published: February 28, 2022

Abstract

In this study, a seed origin discrimination model for *Clinacanthus nutans* was developed. First, 81 *C. nutans* samples from three seed origin locations were collected, and their Near-Infrared (NIR) spectra were obtained. Next, Principal Component Analysis (PCA) was performed on the NIR spectra of the 81 *C. nutans* samples. Then, MSC (multiplicative scatter correction), SNV (standard normal variate), first derivative, and second derivative pre-treatments of the *C. nutans* spectra were performed and combined with the Support Vector Machine (SVM) algorithm for modelling and analysis. Among these methods, first-order derivative pre-treatment achieved the best SVM model effectiveness, with a training set accuracy of 93.44% (57/61) and a test set accuracy of 85.00% (17/20). In order to further improve the discrimination accuracy of the model, three optimization algorithms Grid Search (GS), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) were employed to identify the best c and g parameters for the SVM model. The results demonstrated that the PSO optimization algorithm yielded the best parameters of c = 0.8343, g = 57.8741, with corresponding model training set the accuracy of 96.36% (60/61) and test set the accuracy of 95.00% (20/21). Therefore, developing a seed origin classification model for *C. nutans* based on NIR spectroscopy combined with chemometrics is feasible and has the advantages of being simple, rapid, and green.

Keywords

Near-Infrared Spectroscopy, *Clinacanthus nutans*, Geographical Traceability, SVM, Chemometrics

1. Introduction

Clinacanthus nutans (*C. nutans*), known as the alligator flower, or the Sabah snake grass, is a plant belonging to the genus *Clinacanthus* in the family of *Acanthaceae*. It is found primarily in southern and southwestern China as well as Malaysia, Indonesia, and Thailand [1] [2] [3]. A flurry of research carried out about the chemical composition of *C. nutans*, confirmed that they are rich sources of flavonoids, phenolics, steroids, triterpenoids, cerebrosides, glycoglycerolipids, glycerides, and sulfur-containing glycosides, which make them a useful folk medicine and interesting healthy food [4] [5] [6] [7]. Moreover, various compositional and health studies concluded that *C. nutans* herbal tea has considerable potential as the potential natural antioxidant source. In summary, *C. nutans* may provide beneficial effects on people's health and represent a great economic resource.

As the demand for healthy food growing, consumer attitudes are slowly changing and *C. nutans* are attracting greater interest due to their benefits. Therefore, accurate determination of the origin of *C. nutans* seeds is scientifically important and has application in relevant medicines, health food materials, as well as establishing product quality standards [8].

Currently, the identification of the origin of *C. nutans* seeds and determination of *C. nutans* composition are performed primarily using High-Performance Liquid Chromatography (HPLC) [9] [10] [11] and Gas Chromatography-Mass Spectrometry (GC-MS) [12] [13] [14]. However, high equipment cost, complicated operation, and the need for chemical reagents have restricted their widespread use. Therefore, it is of great significance to develop a rapid, simple, and green method for identifying the origin of *C. nutans* seed.

Near-Infrared (NIR) spectroscopy primarily reveals the overtone bands and combination bands of fundamental vibrations of X-H functional groups (such as C-H, O-H, and N-H) [15] [16] [17] [18]. It not only provides rich qualitative and quantitative information but also is rapid, simple, and does not require chemical reagents. This rapid and simple technique has now been applied in agriculture [19], food science [20], medicine [21], and other fields. Researchers have used NIR spectroscopy combined with chemometrics to confirm the geographical area of durian and have found good application prospects [22]. Herrero Latorre, Peña Crecente, García Martín, and Barciela García [5] used NIR spectroscopy combined with pattern recognition technology to identify honey samples from different sources, developing a fast and single food authentication system to distinguish authentic PGI-Galicia honey samples and other commercial honey samples from other origins. *C. nutans* contains different X-H functional groups with significant absorption in the NIR region. However, there have been few reports

on the geographic origin of *C. nutans* using NIR spectroscopy combined with chemo-metrics.

We collected and analyzed 81 *C. nutans* samples from three geographic locations including Malaysia, Hainan (China), and Guangxi (China). By combining NIR spectroscopy and chemometrics, we established a seed origin classification model for *C. nutans* with high classification accuracy.

2. Materials and Methods

2.1. Experimental Samples

The 81 *C. nutans* samples used in the study originated from Malaysia, Hainan, and Guangxi, China, of which 39 originated from Malaysia, 30 originated from Hainan, and 12 originated from Guangxi. All samples were identified by experts from the Institute of Medicinal Plant Development of Guangdong Academy of Agricultural Sciences.

2.2. Spectral Acquisition

We employed NIRS XDS Rapid Content Analyzer with dispersive grating (FOSS, Denmark) and its diffuse reflectance accessories. The spectrum acquisition range was 400 - 2500 nm, and the detectors were Si (400 - 1100 nm) and PbS (polycrystalline lead sulphide; 1100 - 2500 nm). Spectra were sampled at 2 nm intervals to obtain a range of 400 - 2500 nm. The spectral data of all *C. nutans* samples were collected three times and averaged, and a total of 81 spectra were obtained.

2.3. Sample Set Partitioning

Currently, sample selection methods primarily include the random sampling method, the Kenard-Stone (KS) method, the duplex method, and the sample set partitioning based on joint X-Y distance (SPXY) method. The SPXY method is a sample partitioning method based on the KS method that can be effectively applied to the analysis of the spectral calibration model [23]. Compared to the KS method, the SPXY method considers both the x and y variables when calculating the spatial distance of the sample. The formula for calculating the spatial distance of the x variable is the same as in the KS method (Equation (1)). Equation (2) gives the formula for calculating the spatial distance of the y variable.

$$d_x(p,q)=\sqrt{\sum_{j=1}^{J}\left[x_p(j)-x_q(j)\right]^2};\ p,q\in[1,N] \tag{1}$$

$$d_y(p,q)=\sqrt{\left(y_p-y_q\right)^2}=\left|y_p-y_q\right|;\ p,q\in[1,N] \tag{2}$$

$$d_{xy}(p,q)=\frac{d_x(p,q)}{\max_{p,q\in[1,N]}d_x(p,q)}+\frac{d_y(p,q)}{\max_{p,q\in[1,N]}d_y(p,q)};\ p,q\in[1,N] \tag{3}$$

The stepwise selection process of the SPXY method is similar to that of the KS method, except that $d_{xy}(p,q)$ replaces d_x. $d_{xy}(p,q)$ as the standardized *xy* distance so that the sample has the same weight in *x*- and *y*-spaces. The formula

for this calculation is shown in Equation (3).

In this study, the SPXY method was used to partition the 81 *C. nutans* samples into a training set and a test set at a 3:1 ratio. There were 61 *C. nutans* samples in the training set and 20 *C. nutans* samples in the test set. The details on sample partitioning according to region are shown in Table 1.

2.4. Algorithm

Support vector machine (SVM) is a machine learning method based on statistical learning theory. It has many unique advantages in solving small sample, nonlinear, and high-dimensional pattern recognition problems [24].

The sample training set is represented by $\{(x_i, y_i), i = 1, 2, \cdots, N\}$, where $x_i \in RD$ is the input vector and $y_i \in \{-1, 1\}$ is its corresponding expected output. SVM can identify the optimal hyperplane $(\omega * x) + b = 0$ (where ω is the normal vector of the plane and b is the distance from the plane to the origin) between two categories of data. In cases of linear separability, the data are partitioned into two categories by the plane after classification, and the difference between the two categories of data are $2/\|\omega\|$. The classifier is:

$$f(x) = \text{sgn}\left\{\sum_{i=1}^{N} a_i y_i (x_i \times x) + b\right\} \tag{4}$$

In cases of nonlinearity, SVM maps data from low-dimensional space to high-dimensional space. The classifier is:

$$f(x) = \text{sgn}\left\{\sum_{i=1}^{N} a_i y_i K(x_i \times x) + b\right\} \tag{5}$$

Here, sign{} is the sign function, a_i is a Lagrange multiplier, x_i is a training sample, x is a sample to be classified, and $K(x_i \times x)$ is a kernel function. Selecting the most appropriate kernel function is the most important step in developing a high-performance SVM model, and usually includes two parts: one is to select an appropriate kernel function type, and the other is to optimize the important parameters after determining the kernel function type. Studies have found that models developed with the radial basis function (RBF) kernel selected as the kernel function parameter have good learning ability. Therefore, the RBF kernel function was used in this study to implement SVM modelling. The two important parameters of the RBF kernel function are the penalty parameter c and the kernel function parameter g. These two parameters have significant effects for

Table 1. *Clinacanthus nutans* sample partitioning results.

Origin	Malaysia	Hainan	Guangxi
All sample	39	30	12
Training set	28	23	10
Test set	11	7	2

controlling the complexity, approximation error, and measurement accuracy of the model. Therefore, it is necessary to optimize these two parameters.

Commonly used parameter optimization algorithms include grid search algorithm (GS), genetic search algorithm (GA), and particle swarm optimization algorithm (PSO). GS is a traversal algorithm that tries all (c, g) parameter pairs and then finds the (c, g) parameter pair with the highest accuracy, namely the optimal parameters, through cross-validation [25]. GA is a computational model that simulates natural selection and genetic mechanisms of Darwin's theory of evolution and is a method of searching for an optimal solution [26]. PSO is a stochastic optimization method based on populations. By imitating the swarm behavior of herds, birds, insects, and fish, each member of the group constantly changes its search mode by learning from its and other members' experience [27].

2.5. Model Evaluation Indicators

Model evaluation is used to measure the parameter space and feature extraction effectiveness of different models. The performance of classification models is generally evaluated by the accuracy of the test set [28]. The closer the accuracy is to 1, the better the classification effectiveness of the model. Classification accuracy refers to testing of the established model using the test set in the classification model and is computed as the ratio of the number of statistical samples correctly determined to the total number of samples. In this experiment, the accuracy and the confusion matrix are used for the evaluation of the multi-classification model performance, and the calculation formula is as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{6}$$

$$\text{Confusion matrix} = \begin{bmatrix} \text{TP} & \text{FP} \\ \text{FN} & \text{TN} \end{bmatrix} \tag{7}$$

In the equation, TP represents the number of positive samples from the pre-training set that were correctly classified by the model, FN represents the number of positive samples from the pre-training set that were wrongly classified by the model, FP represents the number of negative samples from the pre-training set that were wrongly classified by the model, and TN represents the number of negative samples from the pre-training set that were correctly classified by the model.

3. Results and Discussion

3.1. Spectral Analysis

C. nutans has a complex composition, including saponins, phenolic compounds, flavonoids, diterpenes, and phytosteroids. These substances have different hydrogen-containing groups and can produce specific absorption bands in the NIR spectrum (780 - 2526 nm), as shown in **Figure 1**. The peaks at 1452 nm and 1939

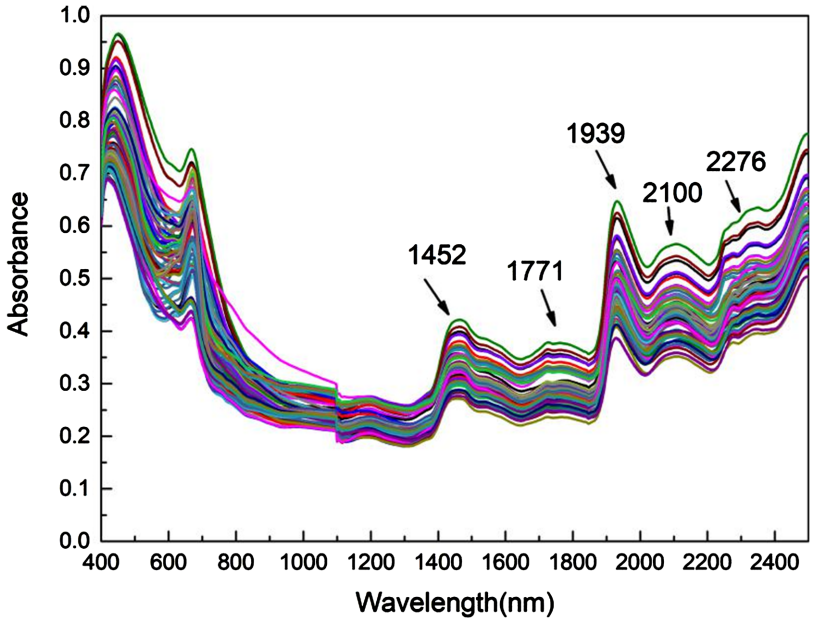

Figure 1. Near-infrared (NIR) spectrum of *Clinacanthus nutans*.

nm are the two major absorption peaks of water in the NIR region. Of these, 1452 nm is the first overtone of O-H stretching vibration, 1939 nm is the combination frequency of O-H stretching and bending vibrations, 1771 nm is the second overtone of C=O absorption, 2100 nm is the O-H deformation vibration and C-O stretching vibration, and 2276 nm is the combination frequency of C-H stretching and deformation vibration [29].

3.2. Principal Component Analysis (PCA)

Due to collinearity between the NIR spectral signals, the information is redundant, as shown in **Figure 1**. The result showed a low difference among the spectral of the 81 samples. Therefore, it is necessary to reduce the dimensionality of the *C. nutans* NIR spectra to simplify the data. PCA is a statistical method for dimensionality reduction using orthogonal transformation to convert the original random vector related to its component into a new random vector whose component is unrelated. This reduces the dimensionality of the multidimensional variable system so that it can be converted into a low-dimensional variable system with high precision (Zou *et al.*, 2006). **Figure 2** represents a PCA score chart of NIR spectrum of *C. nutans*. **Figure 2(a)** represents a two-dimensional score plot for PC1 and PC2. **Figure 2(a)** shows that the samples from the three locations had a wide distribution. Compared to the *C. nutans* samples from Malaysia and Hainan, the samples from Guangxi were more concentrated. **Figure 2(b)** represents a three-dimensional score plot of the first three principal components of *C. nutans* showing the projection of sample points in three-dimensional space. The cumulative total variance obtained by the first three principal components was 95.52%, which indicates that the first three principal components could reflect most of the characteristic information of the original spectrum. The

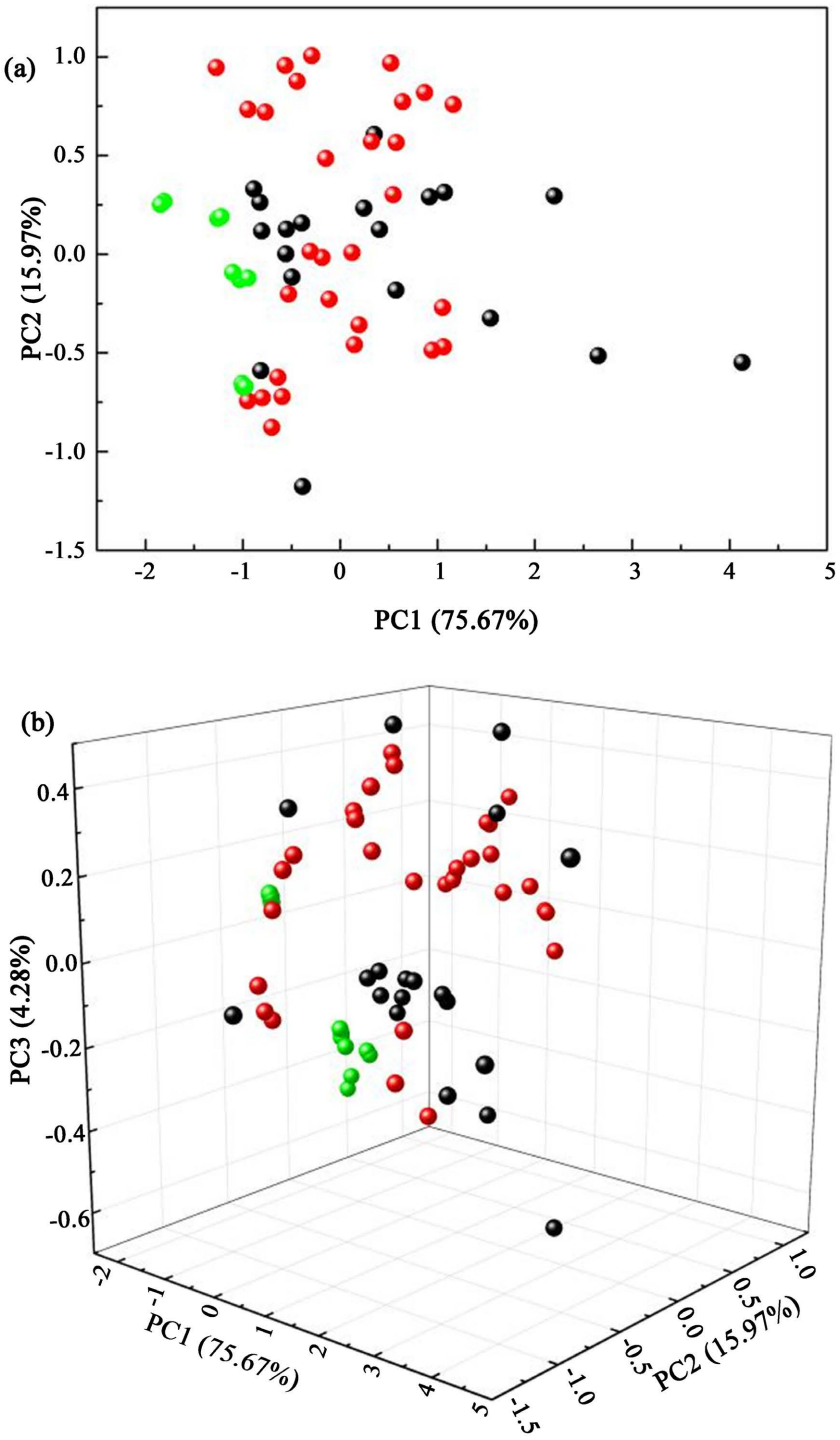

Figure 2. Principal component analysis (PCA) PCA analysis plot of *C. nutans*. (a) 2D analysis plot; (b) 3D analysis plot. Black, Malaysian seed origin; red, Hainan seed origin; green, Guangxi seed origin.

three-dimensional score plot shows that the most dispersed distribution is the *C. nutans* samples from Malaysia, indicating that there is a large intragroup difference in the *C. nutans* samples from Malaysia. The samples from the three *C. nutans* seed locations exhibited large areas of overlap on the PCA score plots. Therefore,

PCA analysis alone cannot be used to make a clear judgment on the origin of *C. nutans* seeds and further algorithmic processing of the *C. nutans* NIR spectra is needed in order to develop a model with high classification accuracy and good prediction accuracy.

3.3. SVM Model Analysis

The SVM has many unique advantages in solving small sample, non-linear, and high-dimensional pattern recognition issues. Thus, the SVM algorithm was used in this study to analyse the NIR spectra of *C. nutans*, and the three parameter optimization algorithms GS, GA, and PSO were used to optimize the two SVM parameters c and g in order to establish a classification model for *C. nutans* seed origin with high accuracy and good predictability.

Data pre-processing is an important factor for improving prediction precision in qualitative analysis and modelling. The acquired spectra not only contains the original information of the samples to be tested but also various external interfering information, which can result in some degree of difference between the measured and true values [30]. In order to eliminate errors as much as possible, various data processing methods must be used to reduce the impact of various interfering factors, thereby laying the foundation for subsequent data processing. In this study, multivariate scattering correction (MSC), standard normal variate transformation (SNV), first derivative, and second derivative were used for pre-treatment of spectral data. **Figure 3** shows the pre-treatment average spectra of the *C. nutans* samples.

In order to compare the effects of different pre-treatment methods on the accuracy of the *C. nutans* seed origin model, SVM models with default c and g parameters (default value of c was 1, default value of g was 1/k, where k was the number of categories) were established for the four pre-treatment methods and compared with the original spectra. The model establishment results in **Table 2** showed that different pre-treatment methods have different effects on the modelling results. Among them, spectra processed by the first derivative yielded the best model prediction effectiveness, with a training set accuracy of 93.44%, and a test set accuracy of 85.00%.

After determining the best pre-treatment method, the parameters c and g were optimized using GS, GA, and PSO. The parameter optimization process and cross-validation results are shown in **Figure 4**. **Figure 4(a)** is a three-dimensional plot of the GS optimization results. As cross-validation accuracy increased, the colour of the grid formed by different c and g values changed from cooler (dark blue) to warmer colour (bright yellow), and at the same time, the horizontal plane of each vertex of the grid increased accordingly. When c = 1 and g = 27.8576, the accuracy of cross-validation reached a maximum of 96.72%. **Figure 4(b)** shows the contour map of the GS parameter optimization results, which is obtained by projecting **Figure 4(a)** onto a two-dimensional plane. **Figure 4(c)** is a plot of GA optimization results. The best-fit curve shows that when

the number of iterations is 0 - 25, the cross-validation accuracy continued to increase. When the number of iterations is 25, the accuracy reached saturation at 96.72%. At this point, c = 1.6327 and g = 55.3856. Figure 4(d) represents a plot

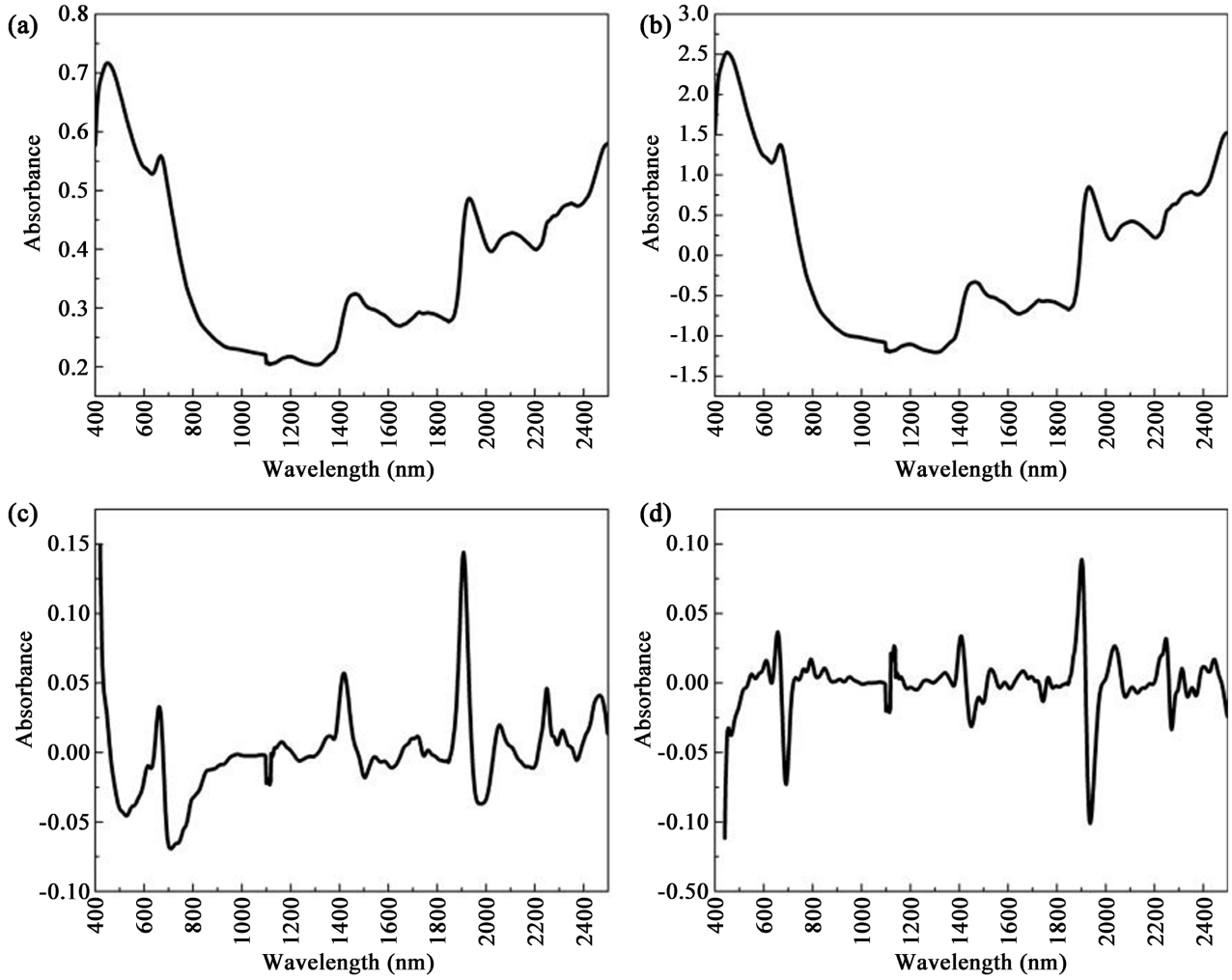

Figure 3. Pre-treated average spectra of *C. nutans* samples. (a) Multivariate scattering correction (MSC); (b) Standard normal variate (SNV); (c) First derivative; (d) Second derivative.

Table 2. Support vector machine (SVM) model classification results using different pre-treatments.

Pre-treatment	Optimization parameters		Training set accuracy	Test set accuracy
	c	g		
Original spectra	default	default	86.88% (53/61)	70.00% (14/20)
MSC	default	default	88.52% (54/61)	75.00% (15/20)
SNV	default	default	90.16% (55/61)	75.00% (15/20)
First derivative	default	default	93.44% (57/61)	85.00% (17/20)
Second derivative	default	default	91.80% (56/61)	80.00% (16/20)

*Multivariate scattering correction (MSC); standard normal variate (SNV)

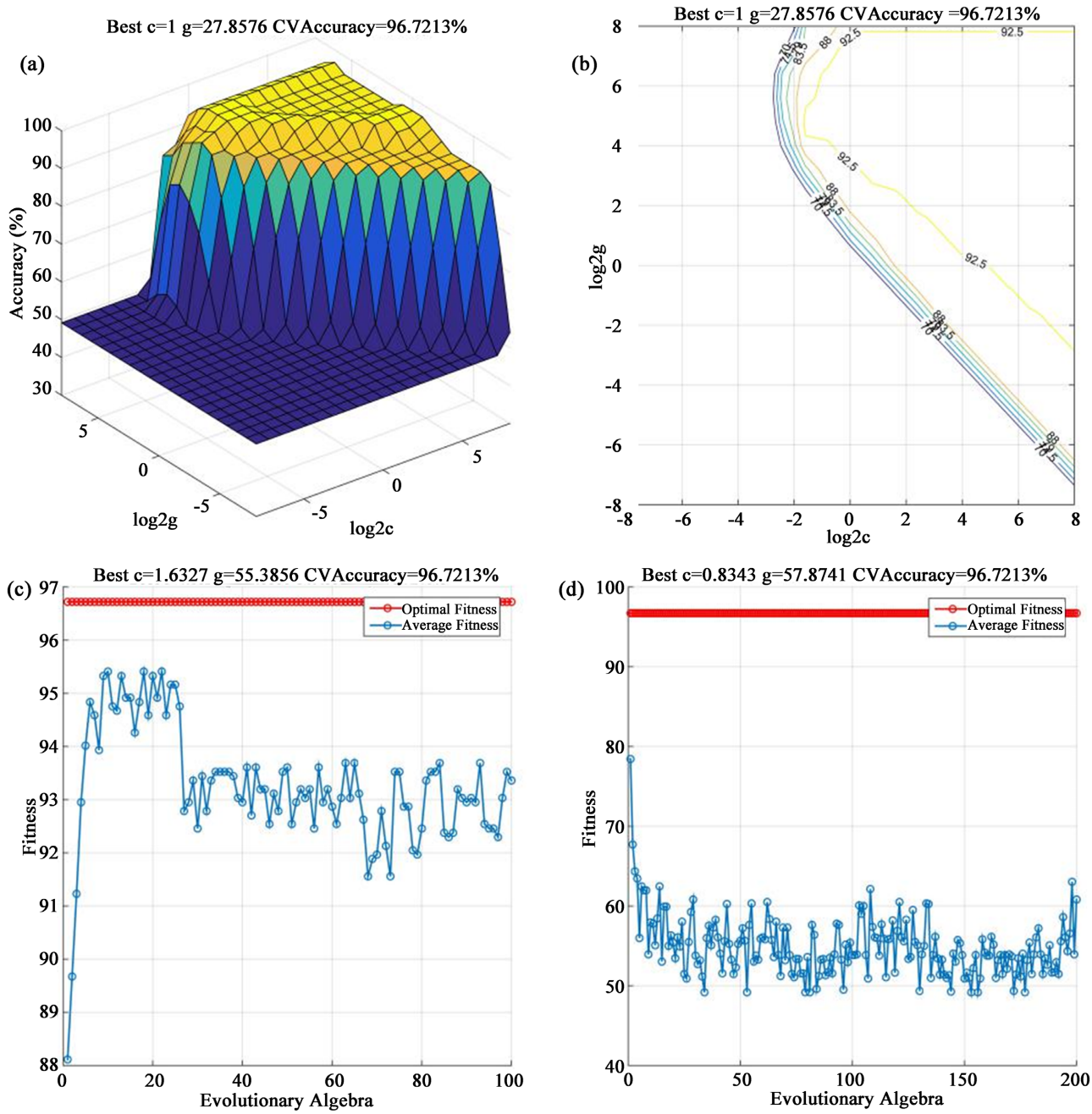

Figure 4. Results of optimizing the parameters c and g in support vector machine (SVM) models. (a) 3D plot of grid search (GS) GS optimization results; (b) Contour plot of GS optimization results; (c) Genetic algorithm (GA) GA optimization results; (d) Particle swarm optimization (PSO) PSO optimization results.

of PSO optimization results. After 50 iterations, the cross-validation accuracy was stable at 97.62%, and the optimal penalty parameter c = 0.8343 and the kernel function parameter g = 57.8741.

After optimizing c and g through the three optimization algorithms GS, GA, and PSO, the cross-validation accuracy reached a minimum of 96.72%. In the next step, the optimal values for c and g were used to establish SVM models and the test set accuracy was used to select the best SVM model. These results are

shown in Table 3. The prediction accuracy of the SVM model was greatly improved after optimization of c and g. The prediction accuracy of the test sets for the three optimization algorithms reached 95.00%, of which PSO yielded the best accuracy. The value of the penalty parameter c was the smallest, therefore, the parameter pair found by PSO was selected as the optimal parameters. The penalty parameter c = 0.8343 and kernel function parameter g = 57.8741 corresponded to the best SVM model for *C. nutans* seed origin, with a training set accuracy of 96.36% (60/61) and a test set accuracy of 95.00% (19/20), the specific results are represented by the confusion matrix in Figure 5.

4. Conclusion

In this study, a classification model for the origin of *C. nutans* seeds based on

Table 3. SVM model classification results after parameter optimization.

Optimization algorithm	Optimization parameters		Cross-validation accuracy	Training set accuracy	Test set accuracy
	c	g			
GS	1	27.8576	96.72%	96.36% (60/61)	95.00% (19/20)
GA	1.6327	55.3856	96.72%	96.36% (60/61)	95.00% (19/20)
PSO	0.8343	57.8741	96.72%	96.36% (60/61)	95.00% (19/20)

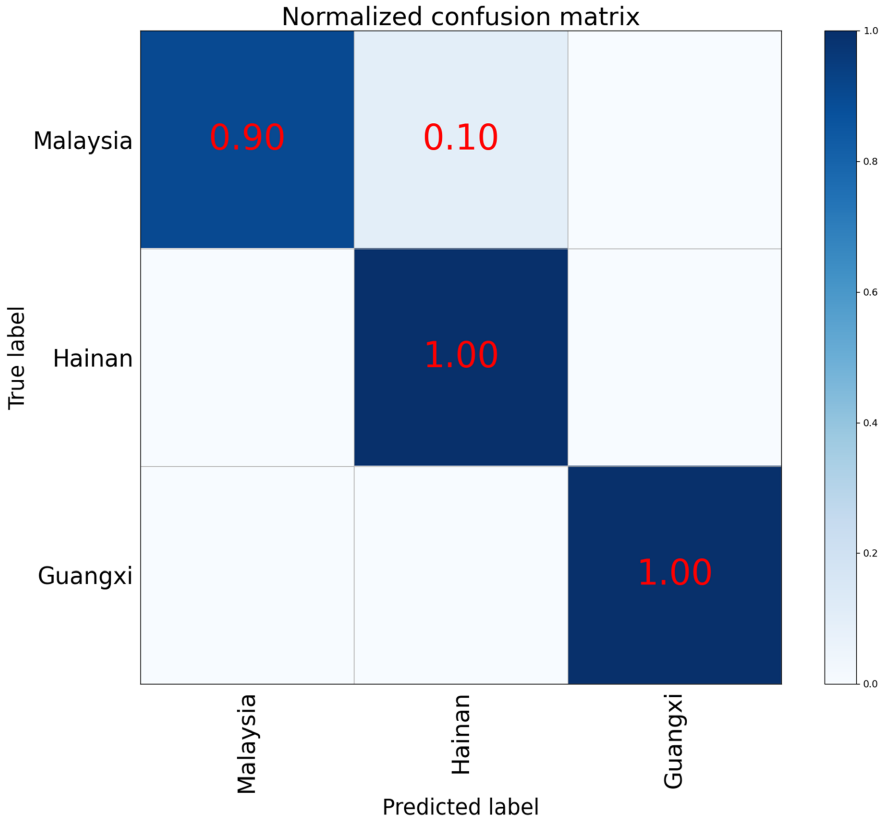

Figure 5. Confusion matrix of SVM model classification results with PSO.

NIR spectroscopy was developed. NIR spectroscopy data were collected from 81 *C. nutans* samples from three geographic locations: Malaysia, Hainan (China), and Guangxi (China). PCA analysis of the acquired NIR spectra showed that the samples from the three geographic locations were dispersed and overlapped in the PCA score plot. Therefore, the NIR spectra of *C. nutans* were further analyzed by SVM modelling. Before the SVM model was established, the spectra were pre-treated with MSC, SNV, the first derivative, and the second derivative. The optimal parameters c and g were found using optimization algorithms. Our results show that the spectra obtained after first derivative processing achieved the best modelling results. The model parameters c = 0.8343 and g = 57.8741 resulted in a training set the accuracy of 96.36% and a test set the accuracy of 95.00%. This method for tracing *C. nutans* seed origin based on NIR spectroscopy combined with chemometrics has the advantages of being simple, rapid, and green.

Acknowledgements

This work was supported by the project funded by China Postdoctoral Science Foundation (2019M663360), the National Natural Science Foundation of China (61975069), the Guangdong Provincial Science and Technology Foundation (2014-a020221068), the Discipline construction project of Guangdong Medical University (4SG21009G), the Funds for PhD researchers of Guangdong Medical University in 2021 (4SG21252G). We also gratefully acknowledge many of our colleagues for their stimulating discussions in this field.

Conflicts of Interest

The authors of the manuscript declare no competing interests.

References

[1] Haida, Z. and Hakiman, M. (2019) A Review of Therapeutic Potentials of *Clinacanthus nutans* as Source for Alternative Medicines. *Sains Malaysiana*, **48**, 2683-2691. https://doi.org/10.17576/jsm-2019-4812-09

[2] Lin, C.M., Chen, H.H., Lung, C.W. and Chen, H.J. (2021) Recent Advancement in Anticancer Activity of *Clinacanthus nutans* (Burm. f.) Lindau. *Evidence-Based Complementary and Alternative Medicine*, **2021**, Article ID: 5560502. https://doi.org/10.1155/2021/5560502

[3] Kamarudin, M.N.A., Sarker, M.M.R., Kadir, H.A. and Ming, L.C. (2017) Ethnopharmacological Uses, Phytochemistry, Biological Activities, and Therapeutic Applications of *Clinacanthus nutans* (Burm. f.) Lindau: A Comprehensive Review. *Journal of Ethnopharmacology*, **206**, 245-266. https://doi.org/10.1016/j.jep.2017.05.007

[4] Zulkipli, I.N., Rajabalaya, R., Idris, A., Sulaiman, N.A. and David, S.R. (2017) *Clinacanthus nutans*: A Review on Ethnomedicinal uses, Chemical Constituents and Pharmacological Properties. *Pharmaceutical Biology*, **55**, 1093-1113. https://doi.org/10.1080/13880209.2017.1288749

[5] Latorre, C.H., Crecente, R.M.P., Martín, S.G. and García, J.B. (2013) A Fast Chemometric Procedure Based on NIR Data for Authentication of Honey with Protected Geographical Indication. *Food Chemistry*, **141**, 3559-3565.

https://doi.org/10.1016/j.foodchem.2013.06.022

[6] Alam, A., Ferdosh, S., Ghafoor, K., Hakim, A., Juraimi, A. S., Khatib, A., *et al.* (2016) *Clinacanthus nutans*: A Review of the Medicinal Uses, Pharmacology and Phytochemistry. *Asian Pacific Journal of Tropical Medicine*, **9**, 402-409. https://doi.org/10.1016/j.apjtm.2016.03.011

[7] Yeo, B.S., Yap, Y.J., Koh, R.Y., Ng, K.Y. and Chye, S.M. (2018) Medicinal Properties of *Clinacanthus nutans*: A Review. *Tropical Journal of Pharmaceutical Research*, **17**, 375-382. https://doi.org/10.4314/tjpr.v17i2.25

[8] de Boer, H.J. and Cotingting, C. (2014) Medicinal Plants for Women's Healthcare in Southeast Asia: A Meta-Analysis of Their Traditional Use, Chemical Constituents, and Pharmacology. *Journal of Ethnopharmacology*, **151**, 747-767. https://doi.org/10.1016/j.jep.2013.11.030

[9] Tan, H.M., Leong, K.H., Song, J., Sufian, N., Hazli, U., Chew, L.Y., *et al.* (2020) Antioxidant and LC-QToF-MS/MS Analysis of Polyphenols in Polar and Non-Polar Extracts from *Strobilanthes crispus* and *Clinacanthus nutans*. *International Food Research Journal*, **27**, 903-914.

[10] Samat, N.M.A.A., Ahmad, S., Awang, Y., Bakar, R.A.H. and Hakiman, M. (2020) Alterations in Herbage Yield, Antioxidant Activities, Phytochemical Contents, and Bioactive Compounds of Sabah Snake Grass (*Clinacanthus nutans* L.) with Regards to Harvesting Age and Harvesting Frequency. *Molecules*, **25**, Article No. 2833. https://doi.org/10.3390/molecules25122833

[11] Zhang, X.H., Zhou, Q., Liu, Z., Qing, X.D., Zheng, J.J., Mu, S.T. and Liu, P.H. (2020) Comparison of Three Second-Order Multivariate Calibration Methods for the Rapid Identification and Quantitative Analysis of Tea Polyphenols in Chinese Teas Using High-Performance Liquid Chromatography. *Journal of Chromatography A*, **1618**, Article ID: 460905. https://doi.org/10.1016/j.chroma.2020.460905

[12] Devasvaran, K., Baharom, N.H., Chong, H.W., Ramli, R.N., Chiu, H.I., Lee, C.K. and Lim, V. (2020) Quality Assessment of *Clinacanthus nutans* Leaf Extracts by GC-MS-Based Metabolomics. *Current Science*, **119**, 641-648.

[13] Ismail, N.Z., Arsad, H., Samian, M.R. and Hamdan, M.R. (2017) Determination of Phenolic and Flavonoid Contents, Antioxidant Activities and GC-MS Analysis of *Clinacanthus nutans* (Acanthaceae) in Different Locations. *Agrivita*, **39**, 335-344. http://doi.org/10.17503/agrivita.v39i3.1076

[14] Alam, M.A., Zaidul, I.S.M., Ghafoor, K., Ferdosh, S., Ali, M.E., Mirhosseini, H., *et al.* (2017) Identification of Bioactive Compounds with GC-Q-TOF-MS in the Extracts from *Clinacanthus nutans* Using Subcritical Carbon Dioxide Extraction. *Separation Science and Technology*, **52**, 852-863. https://doi.org/10.1080/01496395.2016.1271342

[15] Dong, F., Lin, J., You, J., Ji, J., Xu, X., Zhang, L., *et al.* (2020) A Chemometric Modeling-Free Near Infrared Barcode Strategy for Smart Authentication and Geographical origin discrimination of Chinese ginseng. *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy*, **226**, Article ID: 117555. https://doi.org/10.1016/j.saa.2019.117555

[16] Giraudo, A., Grassi, S., Savorani, F., Gavoci, G., Casiraghi, E. and Geobaldo, F. (2019) Determination of the Geographical Origin of Green Coffee Beans Using NIR Spectroscopy and Multivariate Data Analysis. *Food Control*, **99**, 137-145 https://doi.org/10.1016/j.foodcont.2018.12.033 .

[17] Zhou, Y., Zuo, Z., Xu, F. and Wang, Y. (2020) Origin Identification of *Panax notoginseng* by Multi-Sensor Information Fusion Strategy of Infrared Spectra Combined with Random Forest. *Spectrochimica Acta Part A*: *Molecular and Biomolecular Spec-*

troscopy, **226**, Article ID: 117619. https://doi.org/10.1016/j.saa.2019.117619

[18] Ren, G., Ning, J. and Zhang, Z. (2020) Intelligent Assessment of Tea Quality Employing Visible-Near Infrared Spectra Combined with a Hybrid Variable Selection Strategy. *Microchemical Journal*, **157**, Article ID: 105085. https://doi.org/10.1016/j.microc.2020.105085

[19] Neto, A.J.S., Toledo, J.V., Zolnier, S., Lopes, D.D.C., Pires, C.V. and da Silva, T.G.F. (2017) Prediction of Mineral Contents in Sugarcane Cultivated under Saline Conditions Based on Stalk Scanning by Vis/NIR Spectral Reflectance. *Biosystems Engineering*, **156**, 17-26. https://doi.org/10.1016/j.biosystemseng.2017.01.003

[20] Yang, X., Li, Y., Wang, L., Li, L., Guo, L., Huang, F. and Zhao, H. (2019) Determination of 10-Hydroxy-2-Decenoic Acid of Royal Jelly Using Near-Infrared Spectroscopy Combined with Chemometrics. *Journal of Food Science*, **84**, 2458-2466. https://doi.org/10.1111/1750-3841.14748

[21] Liu, Z., Yang, S., Wang, Y. and Zhang, J. (2021) Multi-Platform Integration Based on NIR and UV-Vis Spectroscopies for the Geographical Traceability of the Fruits of Amomum Tsao-Ko. *Spectrochimica Acta Part A*: *Molecular and Biomolecular Spectroscopy*, **258**, Article ID: 119872. https://doi.org/10.1016/j.saa.2021.119872

[22] Chanachot, K., Saechua, W. and Sirisomboon, P. (2021) Geographical Identification of Durian (CV. Monthong) from Pachuap Kiri Khan and Chonburi Province by Using Near Infrared Spectroscopy. *Scientific Paper, Series B*: *Horticulture*, **10**, 67-74.

[23] Tian, H., Zhang, L., Li, M., Wang, Y., Sheng, D., Liu, J., *et al.* (2018) Weighted SPXY Method for Calibration Set Selection for Composition Analysis Based on Near-Infrared Spectroscopy. *Infrared Physics & Technology*, **95**, 88-92. https://doi.org/10.1016/j.infrared.2018.10.030

[24] Wang, X., Esquerre, C., Downey, G., Henihan, L., O'Callaghan, D. and O'Donnell, C. (2021) Development of Chemometric Models Using Vis-NIR and Raman Spectral Data Fusion for Assessment of Infant Formula Storage Temperature and Time. *Innovative Food Science & Emerging Technologies*, **67**, Article ID: 102551. https://doi.org/10.1016/j.ifset.2020.102551

[25] Sun, J., Cong, S.L., Mao, H.P., Zhou, X., Wu, X.H. and Zhang, X.D. (2017) Identification of Eggs from Different Production Systems Based on Hyperspectra and CS-SVM. *British Poultry Science*, **58**, 256-261. https://doi.org/10.1080/00071668.2017.1278625

[26] Feng, Q., Chen, H., Xie, H., Cai, K., Lin, B. and Xu, L. (2020) A Novel Genetic Algorithm-Based Optimization Framework for the Improvement of Near-Infrared Quantitative Calibration Models. *Computational Intelligence and Neuroscience*, **2020**, Article ID: 7686724. https://doi.org/10.1155/2020/7686724

[27] Marini, F. and Walczak, B. (2015) Particle Swarm Optimization (PSO). A Tutorial. *Chemometrics and Intelligent Laboratory Systems*, **149**, 153-165. https://doi.org/10.1016/j.chemolab.2015.08.020

[28] Marjanovic, M., Kovacevic, M., Bajat, B. and Vozenilek, V. (2011) Landslide Susceptibility Assessment Using SVM Machine Learning Algorithm. *Engineering Geology*, **123**, 225-234. https://doi.org/10.1016/j.enggeo.2011.09.006

[29] Li, Y., Shen, Y., Yao, C.L. and Guo, D.A. (2020) Quality Assessment of Herbal Medicines Based on Chemical Fingerprints Combined with Chemometrics Approach: A Review. *Journal of Pharmaceutical and Biomedical Analysis*, **185**, Article ID: 113215. https://doi.org/10.1016/j.jpba.2020.113215

[30] Ma, J.F., Chen, T.L., Zhang, X.D., Xiao, X. and Ge, F.H. (2019) An Approach to Rapid

Determination of Tween-80 for the Quality Control of Traditional Chinese Medicine Injection by Partial Least Squares Regression in Near-Infrared Spectral Modeling. *Journal of Spectroscopy*, **2019**, Article ID: 1521035.
https://doi.org/10.1155/2019/1521035