

Identification of Medicinal Mugua Origin by Near Infrared Spectroscopy Combined with Partial Least-squares Discriminant Analysis

Bangxing Han, Huasheng Peng¹, Hui Yan²

Department of Pharmaceutical Engineering, College of Biological and Pharmaceutical Engineering, West Anhui University, Anhui, Lu'an 237012, ¹Department of Traditional Chinese Medicine Resources, College of Pharmacy, Anhui University of Chinese Medicine, Hefei 230012, ²Department of Biotechnology, School of Biotechnology, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212018, China

Submitted: 24-07-2015

Revised: 04-08-2015

Published: 02-03-2016

ABSTRACT

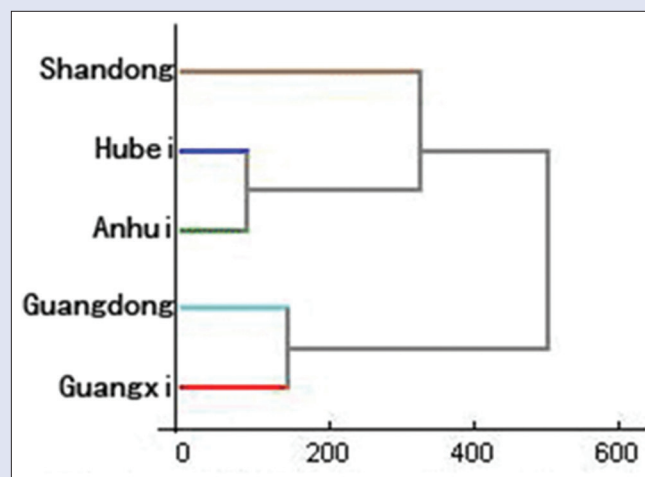
Background: Mugua is a common Chinese herbal medicine. There are three main medicinal origin places in China, Xuancheng City Anhui Province, Qijiang District Chongqing City, Yichang City, Hubei Province, and suitable for food origin places Linyi City Shandong Province. **Objective:** To construct a qualitative analytical method to identify the origin of medicinal Mugua by near infrared spectroscopy (NIRS). **Materials and Methods:** Partial least squares discriminant analysis (PLSDA) model was established after the Mugua derived from five different origins were preprocessed by the original spectrum. Moreover, the hierarchical cluster analysis was performed.

Results: The result showed that PLSDA model was established. According to the relationship of the origins-related important score and wavenumber, and K-mean cluster analysis, the Muguas derived from different origins were effectively identified. **Conclusion:** NIRS technology can quickly and accurately identify the origin of Mugua, provide a new method and technology for the identification of Chinese medicinal materials.

Key words: Chinese geoh herbs, Mugua, near infrared spectroscopy, origin identification, partial least-squares discriminant analysis model

SUMMARY

- After preprocessed by D1+autoscale, more peaks were increased in the preprocessed Mugua in the near infrared spectrum
- Five latent variable scores could reflect the information related to the origin place of Mugua
- Origins of Mugua were well-distinguished according to K-mean value clustering analysis.



Abbreviations used: TCM: Traditional Chinese Medicine, NIRS: Near infrared spectroscopy, SG: Savitzky-Golay smoothness, D1: First derivative, D2: Second derivative, SNV: Standard normal variable transformation, MSC: Multiplicative scatter correction, PLSDA: Partial least squares discriminant analysis, LV: Latent variable, VIP scores: Important score.

Correspondence:

Dr. Hui Yan,
School of Biotechnology, Jiangsu University of
Science and Technology, Sibaidu, Zhenjiang,
Jiangsu 212018, China.
E-mail: yanh1006@163.com
DOI : 10.4103/0973-1296.177907

Access this article online

Website: www.phcog.com

Quick Response Code:



INTRODUCTION

Traditional Chinese Medicine (TCM) is the material basis of TCM clinical application, and its quality is directly related to clinical curative effect. In the long-term practice of TCM, "geoh herbs" becomes the comprehensive evaluation criterion of TCM with excellent quality.^[1,2] The marked feature of geoh herbs is distinct territoriality, so to identify the origin is the important content of study on geoh herbs. The researchers carried out the studies on molecular biology, chemical fingerprint chromatography, various spectroscopy technology, biosensor, and other technologies successively according to the identification of the origin of TCM.^[3] Since 21st Century, computer technology and chemometrics rapidly develop, near infrared spectroscopy (NIRS) is one of the most rapidly developing and widely applied spectral techniques. In recent years, there are many studies on the identification to the origin of TCMs by NIRS.^[4-7] The

spectrum range of NIRS was between 4000 and 12500 cm⁻¹, mainly frequency doubling and frequency harmony absorption of C-H, N-H, and O-H containing hydrogen groups. The sample can be obtained by near infrared spectrometer scanning, including a variety of chemical and physical properties, even biological attribute information. Combined

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

Cite this article as: Han B, Peng H, Yan H. Identification of medicinal Mugua origin by near infrared spectroscopy combined with partial least-squares discriminant analysis. Phcog Mag 2016;12:93-7.

with computer information, chemometrics, artificial intelligence pattern recognition, and other modern technologies, the sample can be analyzed quickly and accurately. It also has simple sample treatment, green environmental protection, no pollution, simultaneous detection of multiple components, and other characteristics, which is widely used in food, TCM, chemical industry, and other fields.^[5-11]

Mugua, a common Chinese herbal medicine, is derived from the dry nearly mature fruit of rosaceous plant *Chaenomeles speciosa* (Sweet) Nakai, which has the functions of calming the liver, relaxing muscles and tendons, harmonizing stomach, and dissipating dampness.^[12] There are three main medicinal origin places in China, Xuancheng City Anhui Province, Qijiang District Chongqing City, and Tujia Autonomous County of Yichang City, Hubei Province. The Muguas from three origins were known as Xuan Mugua, Sichuan Mugua, and Ziqiu Mugua, respectively. Of which Xuan Mugua has always been regarded as geoherb. In recent years, with the development of medicinal Mugua to edible product, the seed resource bred from Linyi City Shandong Province being suitable for food processing was used in the food field.^[13] Therefore, how to identify Xuan Mugua and other Mugua has important significance for clinical medication.

MATERIALS AND METHODS

Instruments

NIRS (Beijing Ruili Analytical Instrument Co., WQF-400N), PbS detector, diffuse reflection loading attachment.

Sample collection and preparation

All the Mugua samples were acquired and identified by Professor Huasheng Peng in Anhui College of TCM, from Guangxi (Y1), Anhui (Y2), Hubei (Y3), Guangdong (Y4), and Shandong Province (Y5), respectively. There were 20 samples from each province, a total of 100 samples. The samples were crushed in advance and sieved with 40 mesh. 10 samples (a total of 50 samples) as calibration set were randomly selected to construct the model. The other 50 samples were used to test the accuracy of the model.

Near infrared data acquisition

Environmental temperature 20°C, relative humidity 45%, scanning range 10,000–3500 cm^{-1} , scanning times 32, resolution 4 cm^{-1} , and light source 10 W/6V halogen tungsten lamp. The air was taken as the control. The spectral data were measured for three times for each sample. The average value was calculated.

Preprocessing of spectra

The medicinal Mugua original spectrum acquired by NIRS contained the relevant sample composition information and a variety of noise signals. The noise signal can produce certain interference to the near infrared spectrum, even affect the calibration model and the prediction of the unknown sample. Therefore, the preprocessing of near infrared spectral data was to solve the effects of various adverse factors on the data information, which laid the basis on the establishment of the calibration model and the accuracy of the prediction set.

The commonly used spectral preprocessing method included savitzky-Golay smoothness, first derivative (D1), second derivative (D2), standard normal variable transformation (SNV), multiplicative scatter correction (MSC), and autoscale. Through the comparison of a variety of preprocessing methods, the particle size, processing environment, and machine noise were investigated. Combined with partial least squares discriminant analysis (PLSDA), the optimum preprocessing methods were optimized, and the optimum preprocessing method was selected.

Modeling of partial least squares discriminant analysis

PLSDA was the regression analysis method of partial least squares algorithm based on discriminant analysis. Similar with the quantitative correction, PLSDA method decomposed spectral array and category array at the same time, highlighted the effect of class information on spectral decomposition, so as to extract the most relevant spectrum information with the sample, namely furthest extract the difference between different spectra. Hence, PLSDA method can usually obtain the better classification and discrimination results than principal component analysis (PCA) and soft independent modeling by class analogy.^[14] It was especially suitable for the situations with the more multiple variables, multicollinearity, small sample size, and bigger influence on all kinds of interference factors.

Data processing

Spectral preprocessing and PLSDA were performed by PLS-toolbox 5.0 (Eigenvector Company USA) software.

RESULTS AND DISCUSSION

Spectral preprocessing results

The original spectrum acquired by the instrument was shown in Figure 1. PLSDA modeling of calibration set was established after the original spectra were preprocessed by D1, D2, SNV, auto scale, and MSC, respectively. The prediction set was used to test the model accuracy. The results showed that D1 + autoscale method was the best, which achieved 100% prediction accuracy in the calibration set (leave-one-out method cross-validation) and prediction set. The spectra were shown in Figure 2 after preprocessed by D1 + autoscale. The comparison between Figures 1 and 2 showed that more peaks were increased in the preprocessed Mugua in the near infrared spectrum, the spectral information was highlighted, which achieved the better preprocessing results.

Partial least squares discriminant analysis modeling

Similar with the PCA and other analysis methods, the near-infrared spectral data were transformed into latent variable (LV) score by PLSDA analysis method. The low LV score can reflect the information contained in the original near infrared spectra, to reduce the dimensionality. The LV cumulative contribution rate in the experiment was shown in Figure 3,

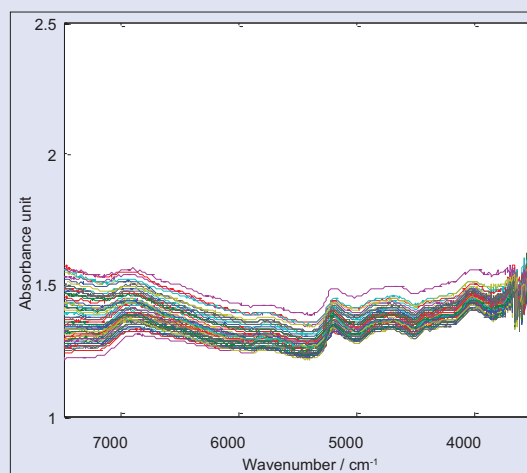


Figure 1: Near-infrared original spectrum of sample

the contribution of the above 3 LV was larger, and the contribution of 4–10 LV was smaller.

All Muguas were classified into three categories by the above 2 LV scores [Figure 4]. All Muguas were classified into four categories by the above 3 LV scores [Figure 5], suggesting that the above 3 LV scores were not enough to completely distinguish five origins of Muguas, the 4th and 5th LV were required.

The optimum accuracy can be achieved when 5 LV scores were used for modeling in the experiment. As shown in Figure 6, the model prediction error rate was decreased with the increase of the LV number. The calibration set and prediction set achieved the best correct rate when 5 LV scores were used, suggesting that the above 5 LV scores could reflect the information related to the origin place.

Latent variable load

The distribution of the above 5 LV loads in different wave number was extracted. Figures 7 and 8 showed that the distribution of the load in the whole wavelength was larger. Hence, for the load

distribution, the spectral information was widely distributed in the whole spectrum.

Important score

Different wave number has great influence on LV score, has an important role to identify the origin of Muguas and is helpful to understand the mechanism of model distinguishing. The relation of the origin-related important score (VIP scores) and wave number was shown in Figure 9, the VIP scores and wavenumber of Muguas from Shandong (Y5) differed from other origin place, 6200–6000 cm^{-1} and 5750–5600 cm^{-1} . The Muguas from other origin places had not obtained the VIP score at the wavenumber.

The VIP score wavenumber in Guangxi (Y1) was similar with that of Guangdong (Y4). Their difference was smaller at 7000–5600 cm^{-1} . For example, VIP score wavenumber was 0 at 6100–6000 cm^{-1} in Guangxi and was negative in Guangdong. The VIP score wave number was negative at 6400–6300 cm^{-1} in Guangxi and was 0 in Guangdong. The VIP score wave number of Muguas in Anhui (Y2) was similar with that of Hubei (Y3). The VIP score wave number of

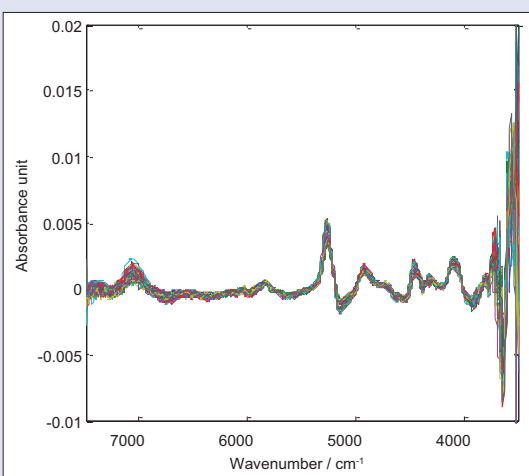


Figure 2: Near-infrared spectrum of sample after preprocessed by D1+autoscale

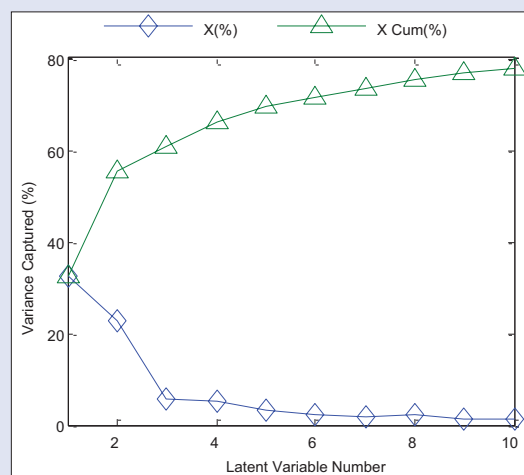


Figure 3: Latent variable cumulative contribution rate

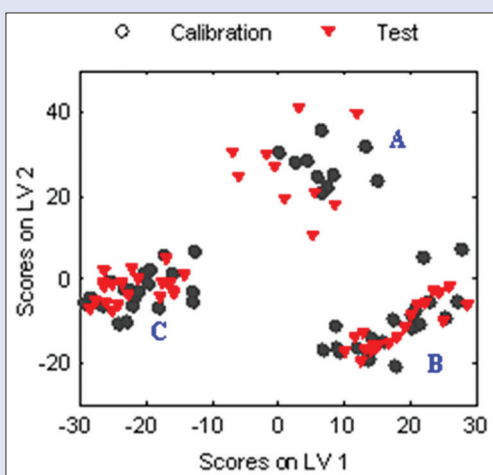


Figure 4: Distribution of the above two latent variable scores. A: Shandong; B: Guangdong; and Guangxi; C: Anhui and Hubei

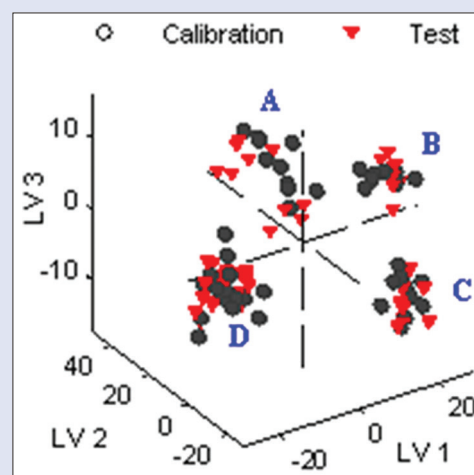


Figure 5: Distribution of the above three latent variable scores. A: Shandong; B: Guangdong; C: Guangxi; D: Hubei and Anhui

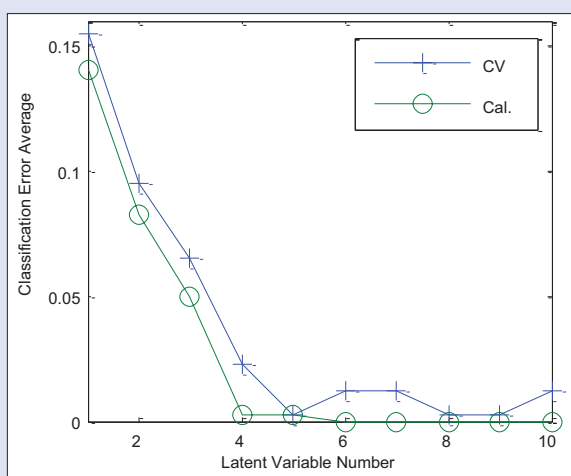


Figure 6: Effect of latent variable score on the model calibration set and prediction set

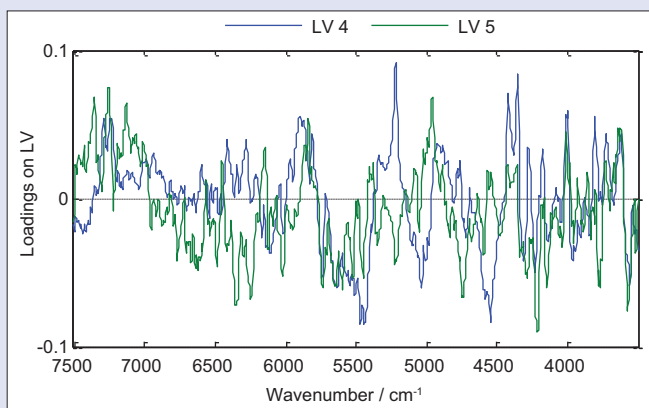


Figure 8: Distribution of the above five latent variable load at different wave number

Muguas in Anhui at 5550 cm^{-1} and 5250 cm^{-1} was higher than that of Hubei.

Different VIP scores may be the basis for the model to differentiate the origins. The scores at different wavenumber were derived from different molecular groups vibration, including different kinds and different quantities, suggesting that the origin had a certain influence on the chemical composition of Mugua. Xuan Mugua VIP score was similar with that of Ziqiu Mugua. They showed the universality in quality. Xuan Mugua at 5550 cm^{-1} and 5250 cm^{-1} was higher than Ziqiu Mugua, thus Xuan Mugua and Ziqiu Mugua can be well-distinguished.

Hierarchical cluster analysis

According to the above 5 LV scores, K-mean value clustering analysis was performed. The results were shown in Figure 10. The distance from Anhui and Hubei Muguas was closest, the distance from Guangdong and Guangxi Muguas was closer. Shandong was far from other origins of Muguas, far away from Guangdong and Guangxi. As for chemical composition, the Muguas in Guangxi was similar with that of Guangdong, the Muguas in Anhui was similar with that of Hubei, but the difference of chemical components was bigger between the two groups of Muguas, they had great difference from Shandong Mugua components. The

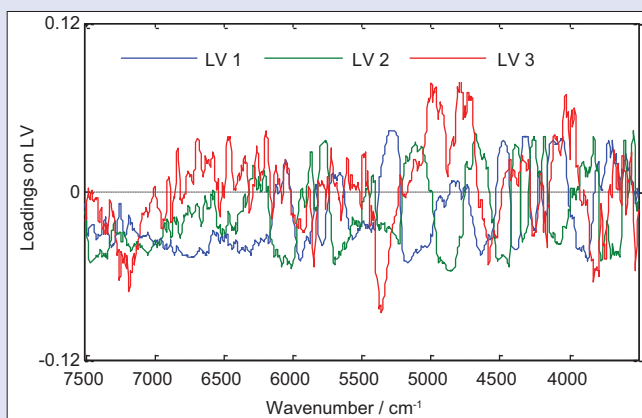


Figure 7: Distribution of the above three latent variable load at different wave number

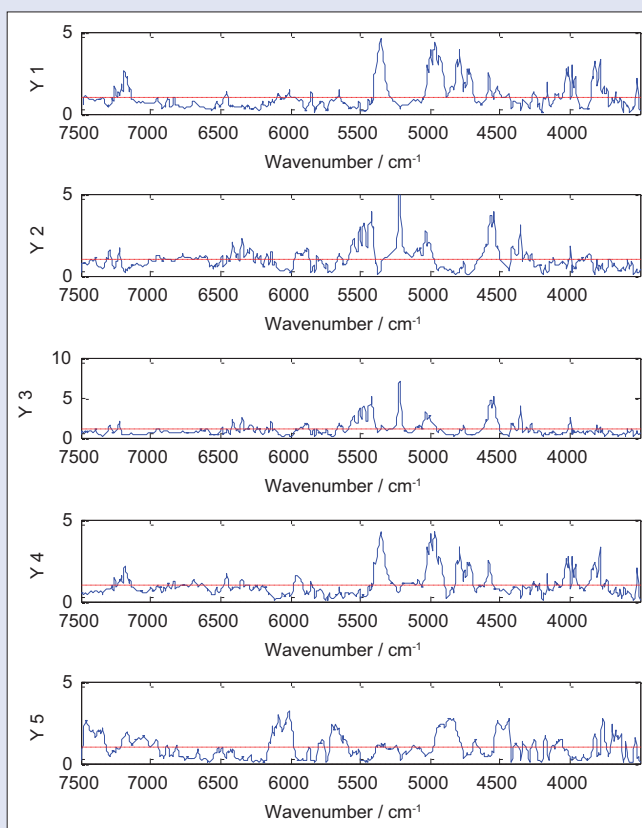


Figure 9: Relationship between important score and wavenumber. Y1: Guangxi; Y2: Anhui; Y3: Hubei; Y4: Guangdong; Y5: Shandong

difference between the Mugua in Shandong, Guangdong, and Guangxi was bigger than those of Anhui and Hubei.

Shandong edible Mugua was bred on the basis of Xuan Mugua introduction. The results of this paper showed that Shandong edible Mugua were clustered with Xuan Mugua and Ziqiu Mugua, showing that the relation among them was closer. They were well-distinguished according to different VIP scores. The spectral VIP scores of Guangxi were similar with those of Guangdong, but it was obviously different from those of three origins.

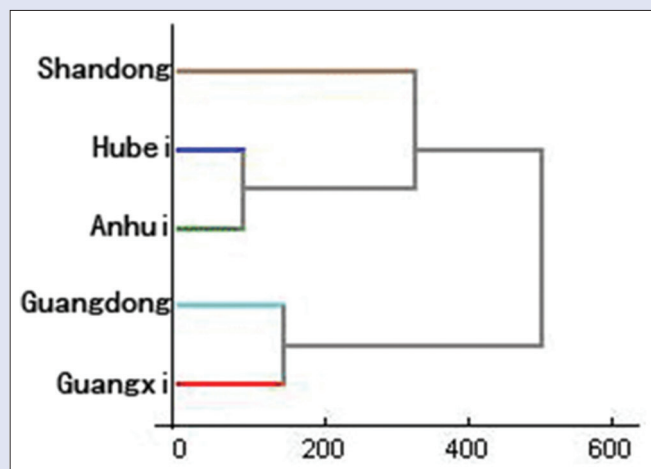


Figure 10: Clustering analysis of Muguas from different origins

CONCLUSION

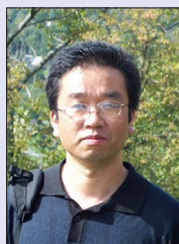
This study provided a fast and nondestructive new method for the identification of Mugua origin through the qualitative analysis of machine learning method combined with NIRS technology. The fast clustering identification was performed Muguas between different origins by infrared spectra fingerprint binding pattern recognition technology. The method is convenient, fast, accurate, suitable for quick identification of a large number of samples, has a certain reliability and practicability. This method provides scientific theory basis for identifying the authenticity of medicinal materials, and quality identification of geoherbs has a broad application prospect.

Financial support and sponsorship

This work was supported by the National Natural Science Foundation of China (Grant No 30901972).

Conflicts of interest

There are no conflicts of interest.



Hui Yan

REFERENCES

- Huang LQ, Guo LP, Hu J, Shao AJ. Molecular mechanism and genetic basis of geoherbs. *Zhongguo Zhong Yao Za Zhi* 2008;33:2303-8.
- Han BX, Peng HS, Huang LQ. Research advances of Dao-di herbs in China. *Chin J Nat* 2012;33:281-5.
- Guo L, Huang L, Huck CW. Near infrared spectroscopy (NIRS) technology and its application in geoherbs. *Zhongguo Zhong Yao Za Zhi* 2009;34:1751-7.
- Han BX, Chen NF, Yao Y. Discrimination of radix *Pseudostellariae* according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition. *Pharmacogn Mag* 2009;5:279-86.
- Han BX, Yan H, Chen CW, Yao HJ, Dai J, Song XW, *et al.* A rapid identification of four medicinal chrysanthemum with near infrared spectroscopy. *Pharmacogn Mag* 2014;10:353-8.
- Yan H, Han BX, Wu QY, Jiang MZ, Gui ZZ. Rapid detection of *Rosa laevigata* polysaccharide content by near-infrared spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc* 2011;79:179-84.
- Blanco M, Alcalá M. Simultaneous quantitation of five active principles in a pharmaceutical preparation: Development and validation of a near infrared spectroscopic method. *Eur J Pharm Sci* 2006;27:280-6.
- Woo YA, Kim HJ, Ze KR, Chung H. Near-infrared (NIR) spectroscopy for the non-destructive and fast determination of geographical origin of *Angelicae gigantis* radix. *J Pharm Biomed Anal* 2005;36:955-9.
- Ghosh SB, Bhattacharya K, Nayak S, Mukherjee P, Salaskar D, Kale SP. Identification of different species of *Bacillus* isolated from nisargruna biogas plant by FTIR, UV-Vis and NIR spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc* 2015;148:420-6.
- Li WL, Han HF, Zhang L, Zhang Y, Qu HB. A feasibility study on the non-invasive analysis of bottled compound E Jiao oral liquid using near infrared spectroscopy. *Sens Actuators B Chem* 2015;211:131-7.
- Li Y, Shi X, Wu Z, Guo M, Xu B, Pan X, *et al.* Near-infrared for on-line determination of quality parameter of *Sophora japonica* L. (formula particles): From lab investigation to pilot-scale extraction process. *Pharmacogn Mag* 2015;11:8-13.
- The Pharmacopoeia Committee of People's Republic of China. Beijing: Chinese Pharmacopoeia; 2010.
- Peng HS, Cheng ME, Wang DQ, Zhang L, Yao Y. Investigation on resource and harvest processing of Mugua. *China J Tradit Chin Med Pharm* 2009;24:1296-8.
- Chu XL, Xu YP, Lu WZ. The study of use of partial least squares in spectroscopy qualitative analysis. *Mod Instrum* 2007;5:13-5.

ABOUT AUTHOR

Hui Yan, is an associate professor at School of Biotechnology, Jiangsu University of Science and Technology, Zhenjiang, China. He research interest is using near-infrared (NIR) spectroscopy to rapid analysis the quality of agricultural products and herbal. Recently, with the development of mobile internet, he is developing the cloud detecting system based on handheld devices and android mobile, through which the handheld NIR devices can be used outside of laboratory by ordinary people.