Pharmacogn. Mag.
A multifaceted peer reviewed journal in the field of Pharmacognosy and Natural Products
www.phcog.com | www.phcog.net

ORIGINAL ARTICLE

Rapid Determination of Puerarin by Near-infrared Spectroscopy During Percolation and Concentration Process of *Puerariae Lobatae Radix*

Xue Jintao[1], Yang Quanwei[2], Jing Yun[1], Liu Yufei[1], Li Chunyan[1,3], Yang Jing[4], Wu Yanfang[1], Li Peng[1], Wan Guangrui[1]

[1]Department of TCM, School of Pharmacy, Xinxiang Medical University, [3]Department of pharmacy, Sanquan Medical College, Xinxiang, [4]Department of pharmacy, Puyang Health School, Puyang, Henan Province, [2]Department of pharmacy, Wu Han NO.1 Hospital, Wuhan, Hubei Province, PR China

Submitted: 16-11-2015 Revised: 30-11-2015 Published: 14-07-2016

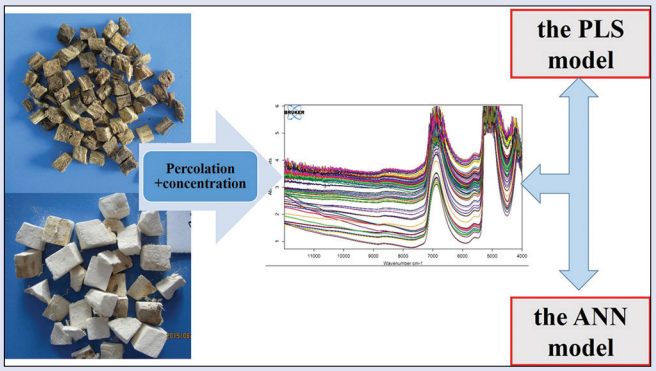
ABSTRACT

Background: Gegen (*Puerariae Labatae Radix*) is one of the important medicines in Traditional Chinese Medicine. The studies showed that Gegen and its preparation had effective actions for atherosclerosis. **Objective:** Near-infrared (NIR) was used to develop a method for rapid determination of puerarin during percolation and concentration process of Gegen. **Materials and Methods:** About ten batches of samples were collected with high-performance liquid chromatography analysis values as reference, calibration models are generated by partial least-squares (PLS) regression as linear regression, and artificial neural networks (ANN) as nonlinear regression. **Results:** The root mean square error of prediction for the PLS and ANN model was 0.0396 and 0.0365 and correlation coefficients (r^2) was 97.79% and 98.47%, respectively. **Conclusions:** The NIR model for the rapid analysis of puerarin can be used for on-line quality control in the percolation and concentration process.

Key words: Artificial neural networks, Gegen (*Puerariae Lobatae Radix*), near-infrared spectroscopy, partial least-squares

SUMMARY

- Near-infrared was used to develop a method for on-line quality control in the percolation and concentration process of Gegen
- Calibration models are generated by partial least-squares (PLS) regression as linear regression and artificial neural networks (ANN) as non-linear regression
- The root mean square error of prediction for the PLS and ANN model was 0.0396 and 0.0365 and correlation coefficients (r^2) was 97.79% and 98.47%, respectively.



Abbreviations used: NIR: Near-Infrared Spectroscopy; Gegen: Puerariae Loabatae Radix; TCM: Traditional Chinese Medicine; PLS: Partial least-squares; ANN: Artificial neural networks; RMSEP: Root mean square error of validation; R2: Correlation coefficients; PAT: Process analytical technology; FDA: The Food and Drug Administration; Rcal: Calibration set; RMSECV: Root mean square errors of cross-validation; RPD: Residual predictive deviation; SLS: Straight Line Subtraction; MLP: Multi-Layer Perceptron; MSE: Mean square error.

Correspondence:

Dr. Wan Guangrui,
School of Pharmacy, Xinxiang Medical University,
Xinxiang, Henan Province, PR China.
E-mail: 591728004@qq.com

DOI: 10.4103/0973-1296.186350

Access this article online
Website: www.phcog.com
Quick Response Code:

INTRODUCTION

Nowadays, high quality of drug products demands on system monitor not only in raw materials, but also in manufacturing processes and processing. The process analytical technology (PAT) initiative by the Food and Drug Administration makes specific recommendations in improving the pharmaceutical manufacturing and quality assurance through process analysis and process control, since it will lead to identification of critical sources of variability, better opportunities to manage process variability, and improved product quality.[1-3]

PAT has also become more and more important in the pharmaceutical industry of Traditional Chinese Medicine (TCM). As is well-known, TCM and its formulae, which have developed a unique system in the past 3000 years, have good effect for different diseases.[4] However, the quality of TCM and its formulae were highly correlated to the quality of manufacturing processes, the crude herbs, and geographical origins (genuineness),[5] so the application of PAT for process analysis was highly desired in food and pharmaceutical industries.

Cite this article as: Jintao X, Quanwei Y, Yun J, Yufei L, Chunyan L, Jing Y, *et al*. Rapid determination of puerarin by near-infrared spectroscopy during percolation and concentration process of *Puerariae Lobatae Radix*. Phcog Mag 2016;12:188-92.

Near-infrared spectroscopy (NIR), associated with chemometrics, has been found to be an effective tool for process analysis.[3] As there are advantages of fast acquisition, no sample preparation, noninvasive, nondestructive, and simultaneous determination of physical properties, and content of several chemical components, NIR spectroscopy with the fiberoptic probes had been widely used in qualitative discrimination and quantitative determination of herb materials, separation monitoring, analysis of extraction process, and so on.[6-10]

Gegen (*Puerariae Lobatae Radix*) is among the earliest and most important medicine used in TCM.[10] Modern studies showed that the main active constituents, puerarin, had been widely used in cardiovascular diseases, acute dysentery, hypertension, alcoholism, deafness, cancer, and diarrhea.[9-12] The previous studies of our group showed that Gegen and its TCM preparation XinMaiJia could effectively inhibit the occurrence and development of atherosclerosis.[13-16]

The objective of this study was to evaluate the possibilities of rapid determination of puerarin during percolation and concentration process of Gegen by NIR. The approach taken was to develop two calibration models by partial least-squares (PLS) as linear regression and artificial neural networks (ANN) as nonlinear regression. The calibration models were optimized individually with high-performance liquid chromatography (HPLC) analysis values as reference. The results of the validation showed the two models were robust, accurate, and repeatable for off-line fast analysis or on-line quality control.

MATERIALS AND METHODS

Samples and reagents

Ten batches of Gegen samples were obtained from different provinces of China. During the percolation and concentration process, 100 samples were collected at different times for HPLC analysis and recording NIR spectral.

Puerarin standards (Batch No. 110752–201313) were obtained from the National Institute for the Control of Pharmaceutical and Biological Products (Beijing, PR China). HPLC grade methanol was purchased from Tedia Company Inc. (Ohio, USA). Water was purified by an ultrapure water instrument. All the other reagents were of analytical grade.

Determination of reference values by high-performance liquid chromatography

As referring to the Chinese Pharmacopoeia (volume II, 2010 edition), the reference values were analyzed immediately after the NIR measurements. A 1260 HPLC system (Agilent Technologies Inc., USA) consisting of ultraviolet–visible detector was used for the quantitative determination of puerarin at a wavelength of 250 nm. A Kromasil C_{18} column (250 mm × 4.6 mm, 5 μm) was employed in 25°C column temperature. The elution system was composed of methanol–water with 0.1% citric acid (25:75) at a flow rate of 1.0 mL/min. The aliquots of 10 μl sample injections were injected into the HPLC for the content analysis.

Near-infrared data collection

As shown in Figure 1, all NIR spectra were recorded by the Bruker Matrix-I FT-NIR spectrometer (Bruker Optik, Ettlingen, Germany) equipped with a PbS detector and a fiberoptic probe. The spectra were obtained over the spectral region of 12000–4000 cm^{-1} at a resolution of 8 cm^{-1} with air absorbance as the reference standard. Each sample measurement was repeated 2 times with 32 scans per spectrum.

Near-infrared data processing

As the most prominent absorption bands of NIR were broad, weak, extensively overlapped bands, so NIR quantification heavily relies on multivariate calibration methods, which were known as linear regression and nonlinear regression. PLS was most frequently used in linear regression, which was depending on a linear relationship between the reference values and the spectral intensity. ANN was most widely used mathematical algorithms for overcoming nonlinearity as the presence of nonlinearity in the spectra, changes of the effective light path, experimental conditions, and instrument variations.[8,17-21]

To develop the best calibration model, the NIR spectroscopic models were constructed by PLS with the OPUS software and ANN with NeuroSolutions for Excel 6.30 (NeuroDimension Inc., Gainesville, FL, USA), respectively.

RESULTS AND DISCUSSION

Development of a linear prediction model by partial least squares

Among the 100 samples, 80 samples were divided randomly into calibration set to establish the calibration model and the remaining 20 samples were into validation set to validate the PLS model. To build the best PLS model, the calibration set must exhibit the full range of potential variations, so when divided into calibration set and validation set, the samples with maximum and minimum contents of puerarin were assigned in the calibration set.

In PLS method, it is well-known that the spectral pretreatment methods, spectral range, and the number of PLS factors are critical parameters for the optimum model, so the best parameter was evaluated and selected based on the correlation coefficients of the calibration set (R_{cal}), the root mean square errors of cross-validation (RMSECV), and the residual

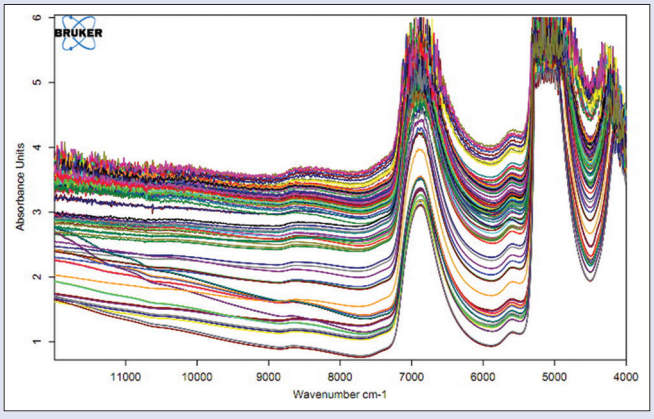


Figure 1: Near-infrared spectra of Gegen samples

Table 1: Parameters of different spectral pretreatment methods

Pretreatment method	R_{cal}^2 (%)	RMSECV	RPD
SLS	98.03	0.0386	7.12
Un-treated	97.48	0.0435	6.30
Constant offset elimination	97.55	0.0429	6.39
First derivative	95.71	0.5530	4.83
Min-max normalization	95.39	0.0589	4.67
First derivative + MSC	94.77	0.0631	4.39
VN	93.79	0.0683	4.05
First derivative + VN	93.50	0.0698	3.94
First derivative + SLS	93.44	0.0708	3.91
MSC	92.12	0.0752	3.56
Second derivative	67.59	0.1510	1.76

SLS: Straight line subtraction; MSC: Multiplicative scatter correction; VN: Vector normalization; RMSECV: Root mean square errors of cross-validation; RPD: Residual predictive deviation

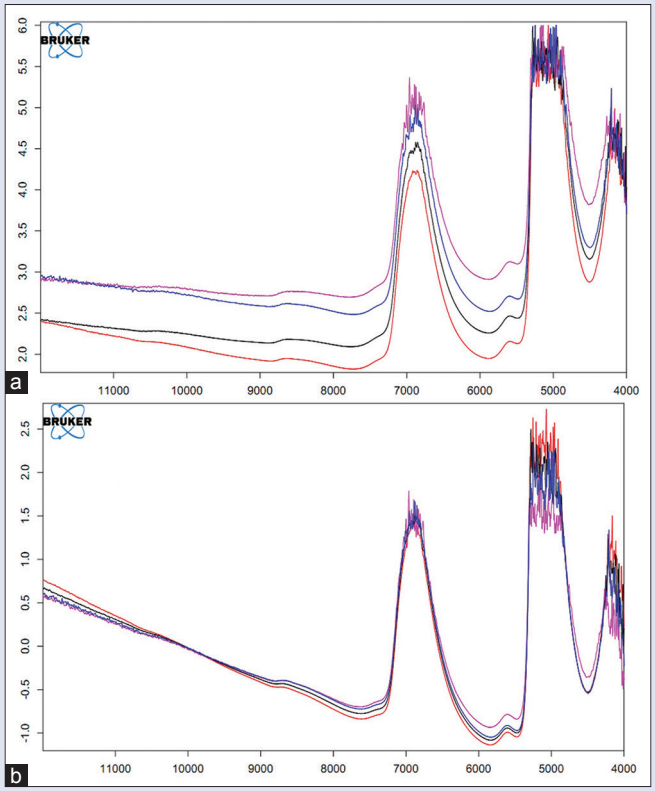


Figure 2: Results of pretreated spectral (a) raw near-infrared spectra, (b) pretreated by straight line subtraction

Table 2: Parameters of different spectral regions

Spectral reign (cm^{-1})	R_{cal}^2 (%)	RMSECV	RPD
11995.2-4246.5	96.79	0.0490	5.58
11995.2-4597.5	98.03	0.0386	7.12
8296.3-4597.5	97.73	0.0413	6.64
11995.2-6097.9	96.36	0.0511	5.25
5453.8-4597.5			

RMSECV: Root mean square errors of cross-validation; RPD: Residual predictive deviation

predictive deviation (RPD), and the model with the best prediction ability was with highest R_{cal} and RPD as well as lowest RMSECV.

According to the above criteria, the calibration models were optimized individually by considering spectral pretreatment methods, spectral range, and number of model factors. As shown in Table 1, 11 spectral pretreatment methods were selected to eliminate or standardize the impact of the spectra and reduce baseline offsets and spectral noise. When compared the parameters of the model, the spectral pretreatment method "straight line subtraction (SLS)" had better performance than other methods. The spectra preprocessed by SLS are presented in Figure 2. Then, different spectral ranges were compared as listed in Table 2, so the best spectral region for the PLS model was in the range from 11995.2 to 4597.5 cm^{-1} which had more efficient information than other spectra range. At last, the number of PLS factors was selected as shown in Figure 3; when number of model factors <7, the RMSECV value declines sharply and the r^2 increases significantly with the increment of PLS factors. However, when more than 7, the value of RMSECV and R2 became gradually, so including more factors in PLS model will not make it set better, but rupture "over-fitting". Therefore, the optimal number of PLS factors was 7.

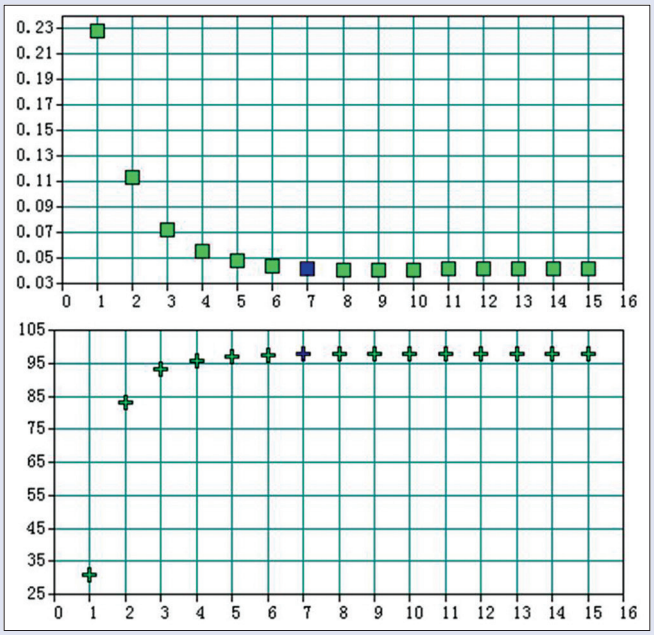


Figure 3: Parameters of different model factors (Up: Root mean square errors of cross-validation factor, Down: r^2-factor)

Development of a nonlinear prediction model by artificial neural networks

Unlike sample assign method of the PLS model, samples in the ANN model were divided into three groups. There were 70 samples in training set, which was used to establish the calibration model, and 20 samples were selected randomly as test set to test the calibration models. The validation set including 43 samples was used to validate the ANN models.

In the study, the ANN model was built based on the multi-layer perceptron network, which was a feed-forward network trained by back propagation. The back-propagation algorithm, the most widely used in ANN, refers to a process which was first performed on an input pattern to calculate the net error from the output to the hidden neurons, then the output error was used to change the weight values in the backward direction, so the error was slowly propagated backward through the hidden layers to minimize the error of the model.[22,23]

The ANN structure was designed individually in spectral pretreatment methods, parameters of network topology, number of hidden neurons, and times of epoch, etc., and the optimum structure was selected in a step-by-step manner based on the evaluation parameters: Higher *R* and lower mean square error (MSE).[22,23]

As shown in Table 3, the parameters of the best ANN model in this work were that vector normalization as spectral pretreatment method, 7 as the number of hidden neurons, TanhAxon function as the transfer function, and 1.1, 0.7, 0.1 and 0.7 as step size and momentum of hidden and output layer neurons, respectively. The times of epoch were 800 as recommended by NeuroSolutions 6.30 to prevent over fitting or early stopping.

Evaluation and validation for the partial least squares and artificial neural networks models

The validation set was used to validate the predictive ability of the optimized PLS model as linear regression model and the optimized

Table 3: The information of the artificial neural networks models

Structure of model	Parameters	MSE	R_{cal}^2 (%)
Spectral pretreatment method	Min max normalization	0.0053	0.9676
	First derivative	0.0058	0.9636
	Vector normalization	0.0024	0.9826
Number of hidden neurons	5	0.0037	0.9765
	5	0.0025	0.9823
	7	0.0024	0.9826
	8	0.0055	0.9695
Transfer function	SigmoidAxon	0.0045	0.9716
	TanhAxon	0.0024	0.9826
	SoftMaxAxon	0.4600	0
	BiasAxon	0.0700	0
	Axon	0.0850	0
Hidden layer			
Step size	0.9	0.0031	0.9798
	1.0	0.0024	0.9826
	1.1	0.0016	0.9901
	1.2	0.0049	0.9682
Momentum	0.55	0.0030	0.9817
	0.60	0.0016	0.9898
	0.70	0.0016	0.9901
	0.80	0.0035	0.9855
Output layer			
Step size	0.05	0.0055	0.9617
	0.10	0.0016	0.9901
	0.15	0.0022	0.9859
Momentum	0.6	0.0072	0.9587
	0.7	0.0016	0.9901
	0.8	0.0031	0.9795

MSE: Mean square error

ANN model as nonlinear regression model. The results showed than the root MSE of prediction (RMSEP) for the PLS and ANN model was 0.0396 and 0.0365 and correlation coefficients (r^2) was 97.79% and 98.47%, respectively. The T and F tests indicated there was no significant difference between the predicted result and the reference value, so the established two models were robust, accurate, and repeatable for monitoring the percolation and concentration process.

Compared to the PLS model, both RMSEP and r^2 in the ANN models were much better, so the ANN model had more advantages for the rapid analysis of puerarin and routine screening in the quality control of the XinMaiJia preparation.

CONCLUSIONS

In this study, NIR spectroscopy was developed for the rapid analysis of puerarin in the percolation and concentration process of Gegen. By means of PLS and ANN regression, robust NIR models were built with high correlations and low error of predictions. Compared with the HPLC methods, NIR realized fast acquisition, no sample preparation, and nondestructive, so it can significantly save manpower and time. As a PAT tool, NIR could be carried out in off-line fast analysis or on-line quality control in the quality control of the XinMaiJia preparation, and this method will significantly improve the efficiency of quality control in TCM.

Acknowledgment

Xue Jintao and Yang Quanwei contributed equally to this work and should be considered as co-first authors. We are extremely grateful for the financial support provided by the National Sciences Foundation of PR China (No. 21505114), the Special Development Fund of Henan Provincial Department of Finance (No. Yu[2013] 398–19), the Key Project of Science and Technology of Henan Provincial Department of Science and Technology (No. 132102310247) and the Cultivation Fund of Xinxiang Medical University (No. 505095). The authors would like to acknowledge the collaborators from Huasun Group Inc., Ltd., for supporting this work.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Knop K, Kleinebudde P. PAT-tools for process control in pharmaceutical film coating applications. Int J Pharm 2013;457:527-36.
2. Krier F, Mantanus J, Sacré PY, Chavez PF, Thiry J, Pestieau A, *et al.* PAT tools for the control of co-extrusion implants manufacturing process. Int J Pharm 2013;458:15-24.
3. Stenlund H, Johansson E, Gottfries J, Trygg J. Unlocking interpretation in near infrared multivariate calibrations by orthogonal partial least squares. Anal Chem 2009;81:203-9.
4. Liu J, Zhang H, Ji B, Cai S, Wang R, Zhou F, *et al.* A diet formula of Puerariae radix, *Lycium barbarum, Crataegus pinnatifida*, and Polygonati rhizoma alleviates insulin resistance and hepatic steatosis in CD-1 mice and HepG2 cells. Food Funct 2014;5:1038-49.
5. Yu K, Cheng YY. Discriminating the genuineness of Chinese medicines with least squares support vector machines. Chin J Anal Chem 2006;34:561-4.
6. Andersson M, Folestad S, Gottfries J, Johansson MO, Josefson M, Wahlund KG. Quantitative analysis of film coating in a fluidized bed process by in-line NIR spectrometry and multivariate batch calibration. Anal Chem 2000;72:2099-108.
7. Xue JT, Wu CJ, Wang LL, Zhang JL, Huang G, Jiang S, *et al.* Identification of the types of parched areca seed by near infrared spectroscopy. Asian J Chem 2012;24:822-6.
8. Xue J, Wu C, Wang L, Jiang S, Huang G, Zhang J, *et al.* Dynamic prediction models for alkaloid content using NIR technology for the study and online analysis of parching in *Areca* seed. Food Chem 2011;126:725-30.
9. Wang P, Zhang H, Yang H, Nie L, Zang H. Rapid determination of major bioactive isoflavonoid compounds during the extraction process of kudzu (*Pueraria lobata*) by near-infrared transmission spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 2015;137:1403-8.
10. Lau CC, Chan CO, Chau FT, Mok DK. Rapid analysis of radix puerariae by near-infrared spectroscopy. J Chromatogr A 2009;1216:2130-5.
11. Lau C, Carrier D, Beitle R, Howard L, Lay J, Liyanage R, *et al.* A glycoside flavonoid in kudzu (*Pueraria lobata*): Identification, quantification, and determination of antioxidant activity. Appl Biochem Biotechnol 2005;123:783-94.
12. Jin SE, Son YK, Min BS, Jung HA, Choi JS. Anti-inflammatory and antioxidant activities of constituents isolated from Pueraria lobata roots. Arch Pharm Res 2012;35:823-37.
13. Yin Y, Wan J, Li P, Jia Y, Sun R, Pan G, *et al.* Protective effect of Xin Mai Jia ultrafiltration extract on human umbilical vein endothelial cell injury induced by hydrogen peroxide and the effect on the NO-cGMP signaling pathway. Exp Ther Med 2014;8:38-48.
14. Wan J, Yin Y, Sun R, Pan G, Li P, Jia Y, *et al.* Protective effect of the ultra-filtration extract from Xin Mai Jia on human aortic smooth muscle cell injury induced by hydrogen peroxide. Exp Ther Med 2014;7:11-16.
15. Li P, Pan GP, Jia M, Wang QQ, Guo ZG, Zhao FR, *et al.* Effect of Xin Mai Jia on atherosclerosis in rats. Genet Mol Res 2015;14:6018-27.
16. Yin YL, Li P, Yang J, Liu DW, Sun RL, Pan GP, *et al.* Protective effect of ultrafiltered Xin Mai Jia extract against H2O2-induced injury in human umbilical vein endothelial cells through NHE1 downregulation. Genet Mol Res 2014;13:8436-49.
17. Şenyiğit E, Atici U. Artificial neural network models for lot-sizing problem: A case study. Neural Comput Appl 2013;6:1039-47.
18. Blanco M, Coello J, Iturriaga H, Maspoch S, Pages J. NIR calibration in non-linear systems: Different PLS approaches and artificial neural networks. Chemometr Intell Lab 2000;50:75-82.
19. Falode O, Sunday O. Predicting the onset of asphaltene precipitation in heavy crude oil using artificial neural network. Chem Process Eng Res 2013;15:1-10.
20. Heigl N, Greiderer A, Petter CH, Kolomiets O, Siesler HW, Ulbricht M, *et al.* Simultaneous determination of the micro-, meso-, and macropore size fractions of porous polymers by

a combined use of Fourier transform near-infrared diffuse reflection spectroscopy and multivariate techniques. Anal Chem 2008;80:8493-500.

21. Yu HY, Niu XY, Lin HJ, Ying YB, Li BB, Pan XX. A feasibility study on on-line determination of rice wine composition by Vis-NIR spectroscopy and least-squares support vector machines. Food Chem 2009;113:291-6.
22. Rizkalla N, Hildgen P. Artificial neural networks: Comparison of two programs for modeling a process of nanoparticle preparation. Drug Dev Ind Pharm 2005;31:1019-33.
23. Mirsepahi A, Chenb L, O'Neilla B. A comparative approach of inverse modelling applied to an irradiative batch dryer employing several artificial neural networks. Int commun heat mass 2014;53:164-73.

Wan Guangrui

ABOUT AUTHOR

Wan Guangrui, is a Professor at the School of Pharmacy, Xinxiang Medical University, Xinxiang, Henan Province, PR China. His research interest is in the area of pharmacodynamic study of the active ingredients of Traditional Chinese Medicine (TCM). In the past years, he mainly studied the TCM preparation XinMaiJia that effectively inhibit the occurrence and development of atherosclerosis.