Journal of Innovative Optical Health Sciences
Vol. 7, No. 6 (2014) 1450011 (9 pages)

DOI: 10.1142/S1793545814500114

World Scientific
www.worldscientific.com

Application of particle swarm optimization-based least square support vector machine in quantitative analysis of extraction solution of yangxinshi tablet using near infrared spectroscopy

Weijian Lou*, Kai Yang†, Miaoqin Zhu‡, Yongjiang Wu†,
Xuesong Liu† and Ye Jin†,§
*Department of Pharmacy
Sir Run Run Shaw Hospital of School of Medicine
Zhejiang University, Hangzhou 310016, P. R. China
†College of Pharmaceutical Sciences
Zhejiang University, Hangzhou 310058, P. R. China
‡Department of Chemistry
Zhejiang International Studies University
Hangzhou 310012, P. R. China
§iloveleafking@zju.edu.cn

Received 19 July 2013
Accepted 10 November 2013
Published 2 January 2014

A particle swarm optimization (PSO)-based least square support vector machine (LS-SVM) method was investigated for quantitative analysis of extraction solution of Yangxinshi tablet using near infrared (NIR) spectroscopy. The usable spectral region (5400–6200 cm^{-1}) was identified, then the first derivative spectra smoothed using a Savitzky–Golay filter were employed to establish calibration models. The PSO algorithm was applied to select the LS-SVM hyper-parameters (including the regularization and kernel parameters). The calibration models of total flavonoids, puerarin, salvianolic acid B and icariin were established using the optimum hyper-parameters of LS-SVM. The performance of LS-SVM models were compared with partial least squares (PLS) regression, feed-forward back-propagation network (BPANN) and support vector machine (SVM). Experimental results showed that both the calibration results and prediction accuracy of the PSO-based LS-SVM method were superior to PLS, BP-ANN and SVM. For PSO-based LS-SVM models, the determination coefficients (R^2) for the calibration set were above 0.9881, and the RSEP values were controlled within 5.772%. For the validation set, the RMSEP values were close to RMSEC and less than 0.042, the RSEP values were under 8.778%, which

were much lower than the PLS, BP-ANN and SVM models. The PSO-based LS-SVM algorithm employed in this study exhibited excellent calibration performance and prediction accuracy, which has definite practice significance and application value.

Keywords: Near infrared spectroscopy; extraction; particle swarm optimization; least square support vector machines.

1. Introduction

Yangxinshi tablet is a traditional Chinese medicine (TCM) tablet, which is used clinically to treat coronary disease, angina pectoris, myocardial infarction, hyperlipidemia and hyperglycemia.[1] The preparation of Yangxinshi tablet consists of 13 medicine herbs: Radix Astragali, Radix Codonopsis, Radix Salviae Miltiorrhizae, Radix Puerariae, Folium Epimedii, Radix Rehmanniae, Radix Angelicae Sinensis, Ganoderma Lucidum, Radix Glycyrrhiza, etc. Extraction is the most initial manufacturing process for Yangxinshi tablet, directly relevant to the utilization of crude materials. Currently, the quality control of extraction process relies heavily on stipulated extraction time and high performance liquid chromatography (HPLC).[2,3] Although reliable and relatively accurate, HPLC is time consuming and requires complex sample preparation, limiting its application to rapid detection and process control. Therefore, a fast and accurate method is required to speed up the determination of the intermediate quality attributes to permit quality control during pharmaceutical production.

Near infrared (NIR) spectroscopy is a simple, fast, and nondestructive technique; it enables the analysis of samples without complicated pretreatments, which results in substantially decreased analysis time relative to traditional analytical methods, e.g., chromatographic techniques.[4] NIR spectroscopy is mainly used to record information in the overtone and combination band regions of the spectrum.[5] Direct quantification analysis based on the complexity and high dimension of NIR spectral data is difficult. Calibration methods, such as PLS, LS-SVM and artificial neural networks (ANN), are in more common use at present.

In this study, PSO-based LS-SVM regression[6–8] method was employed to extract the NIR spectral features and to investigate the correlation between the spectral data and the concentration variables measured by the reference assays. PSO was used to optimize the hyperparameters of the LS-SVM model, including the regularization and kernel parameters. LS-SVM has the capability for linear and nonlinear multivariate calibration.[9] The LS-SVM models were developed for quantitative analysis of the critical intermediate quality attributes of interest: total flavonoids, puerarin, salvianolic acid B and icariin. Then, the calibration performance and prediction accuracy of PSO-based LS-SVM were generally compared with the conventional PLS, feed-forward back-propagation network (BPANN) and SVM methods.

2. Materials and Methods

2.1. *Materials*

Crude materials (Radix Astragali, Radix Codonopsis, Radix Salviae Miltiorrhizae, Radix Puerariae, Folium Epimedii, Radix Rehmanniae, Radix Angelicae Sinensis, Ganoderma Lucidum and Radix Glycyrrhiza) were provided by the Qingdao Growful Pharmaceutical Co., Ltd. The standards of puerarin, salvianolic acid B and icariin (over 99% purity) were purchased from the National Institutes for Food and Drug Control (Beijing, China). Methanol, chromatographically pure grade, was purchased from Merck (Darmstadt, Germany) for HPLC analysis. Water was purified by Millipore water purification device (Millipore Corp., Billerica, MA, USA). All other reagents were of analytical grade.

2.2. *Extraction process and sampling*

Mixed extraction of Yangxinshi tablet was carried out in a 5-L extractor. Suitable amounts of Radix Astragali, Radix Codonopsis, Radix Salviae Miltiorrhizae, Radix Puerariae, Folium Epimedii, Radix Rehmanniae, Radix Angelicae Sinensis, Ganoderma Lucidum and Radix Glycyrrhiza were extracted with 1.6 L of distilled water for 2 h in the

first extraction process, and with 1.2 L of distilled water for 1.5 h in the second extraction process. Extract solutions were withdrawn from each extraction process, at regular intervals, and analyzed for the same parameters by reference assays. In order to obtain similar prediction accuracy, it was necessary to ensure uniform distribution of samples with high or low concentrations.[10] Therefore, samples were collected at 8 min interval in the first 40 min of the first extraction process, at 10 min interval in the last time of the first extraction process, and at 10 min interval in the second extraction process.

In this study, samples were collected from extraction processes in six batches and 144 samples were obtained in total, which were used as a calibration set for the model development. Further extraction processes providing new data were performed for the external validation of the models.

2.3. *Spectral measurement*

Near infrared transmittance spectra of samples from the extraction processes were collected at 8 cm^{-1} interval over the spectral range of 4000–10000 cm^{-1} using an Antaris (Thermo Fisher Scientific Inc., Madison, WI, USA) FT-NIR spectrometer. Each sample was scanned with a 2-mm path length and a spectrum obtained by averaging 32 scans. All spectra were collected in absorbance mode. The average spectrum from each triplicate measurement as the final spectrum of each sample was used to quantify acteoside concentration. Data acquisition and spectral pretreatments were performed using the RESULT^{TM} software suite (version 3.0, Thermo Nicolet, USA), TQ analyst (version 8.0, Thermo Nicolet, USA) and Matlab (version 7.5, The MathWorks, Inc., Natick, MA, USA).

2.4. *Reference assays*

A high-performance liquid chromatographic (HPLC) method was developed to quantitatively determine puerarin, salvianolic acid B and icariin in the extract solution samples, and the results were used as reference data for the NIR analysis. Chromatographic analysis was carried out on an Agilent 1200 HPLC system (Agilent Technologies, USA) equipped with a vacuum degasser, a quaternary pump, an autosampler, a thermostatic column compartment and a diode array detector (DAD). Separation was performed on an Agilent Eclipse XDB-C18 column (250 mm × 4.6 mm, 5 μm) at 30°C. The mobile phase was methanol (A) and 0.3% phosphoric acid aqueous solution (B) with a gradient program as follows: 0–15 min, 25% A; 15–17 min, 25–40% A; 17–30 min, 40–45% A; 30–31 min, 45–55% A; 31–43 min, 55–60% A; and 43–45 min, 60–100% A. The flow rate of mobile phase was maintained at 1 mL/min. The quantitative chromatograms in this study were collected at 250 nm in the former 23 min, at 286 nm between 23 and 33 min, and at 270 nm until the end of analysis. Each run was followed by equilibration time of 5 min. The samples collected from the extraction processes were centrifuged at 13,000 rpm for 10 min, and the supernatant fluid was passed through a syringe filter of 0.45 μm and 5 μL of the filtrate was injected into the HPLC system for analysis. The chemical structures of puerarin, salvianolic acid B and icariin and a typical HPLC chromatogram are shown in Fig. 1.

The concentration of total flavonoids was determined by ultraviolet and visible spectrophotometry.[11] Stock solution of puerarin was prepared by dissolving the accurately weighed puerarin standard in methanol, and the concentration was 0.2 mg/mL. Then, the stock solution was diluted to appropriate concentration range to establish a calibration curve. The extraction solutions collected

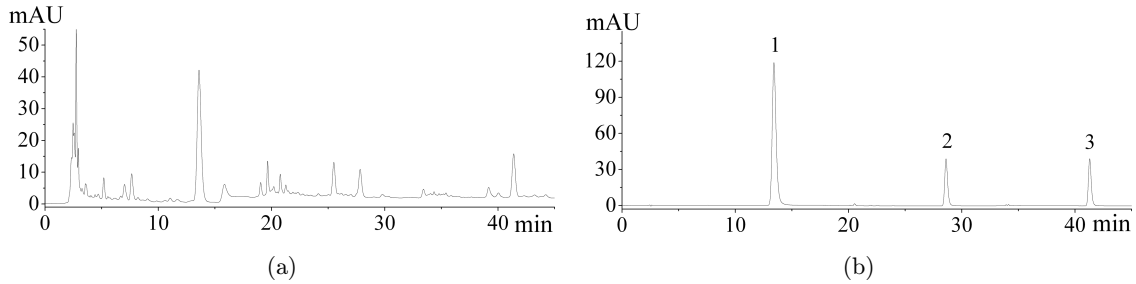

Fig. 1. Typical HPLC chromatograms of an extraction solution (a) and the mixed standard solution (b) (1. Puerarin, 2. salvianolic acid B, 3. icariin).

during the first and second extraction processes were centrifuged at 13,000 rpm for 10 min. Then, the supernatant fluid of 1 mL was diluted to 125–250 mL (according to the actual concentration) with methanol. Absorption intensities were measured at 250 nm using methanol as blank. The concentration of total flavonoids was calculated from the calibration curve.

2.5. *Method validation for HPLC*

The calibration curves were constructed by plotting the peak area (y) vs the concentration (x, μg/mL) of each analyte. The linear regression and the determination coefficient (R^2) were as follows: $y = 14.11x - 13.16 (R^2 = 0.9999)$ for puerarin; $y = 3.138x - 1.291 (R^2 = 0.9999)$ for salvianolic acid B; $y = 6.313x - 1.068 (R^2 = 0.9999)$ for icariin. Good linearity was achieved in the investigated ranges. The linear ranges were 51.5–515, 56.2–562 and 27.6–276 μg/mL for puerarin, salvianolic acid B and icariin, respectively.

The repeatability of the method was determined by analyzing six independently prepared solutions of the first sample collected from the first extraction process of batch 1 (see Fig. 1). The relative standard derivation (RSD) values of the peak areas for the three analytes were found to be 0.03–1.13%.

The stability was tested at room temperature by analyzing the same extraction solution every 4 h for 24 h. The RSD values of the peak areas for the three analytes were all less than 1.72%, indicating that the sample solutions were stable at least for 24 h.

The percent recovery was determined by adding accurately known amounts of the corresponding analytes at three concentration levels (0.5, 1.0 and 1.5 times the concentration in the matrix) to the previously analyzed sample and reanalyzing the sample. The percent recovery was between 94.9% and 103.3%, with RSD values less than 3.31% for all three analytes.

The above validation results indicated that the developed analytical method used for quantitative determination was acceptable. Typical HPLC chromatograms of an extraction solution (A) and the mixed standard solution are shown in Fig. 1.

2.6. *Chemometrics and data analysis*

In this study, PSO-based LS-SVM, PLS, BP-ANN and SVM as four linear and nonlinear regression methods were investigated to build quantitative models for rapid determination of extraction solutions of Yangxinshi tablet. The performance and the predictive ability of the established models were assessed in terms of R^2 (the determination coefficient), RMSEC (the root mean square error of calibration), RMSEP (the root mean square error of prediction) and RSEP (the relative standard error of prediction).[12,13]

SVM,[14] which is based on the idea of structural risk minimization (SRM), was initially developed by Vapnik as a binary classification tool.[15] The principles of SVM can easily be extended to regression tasks. In the SVM regression approach, input vectors are mapped into a high-dimensional feature space where linear regression is performed. The problem is formulated as constrained quadratic optimization in the high-dimensional space.

As a reformulation of standard SVM, LS-SVM was introduced by Suykens *et al.* and leads to solving the linear Karush–Kuhn–Tucker (KKT) systems for classification and regression problems.[14] In the LS-SVM algorithm, the parameters for regularization and in the kernel function, the so-called hyperparameters, play a crucial role in the algorithm performance. Radial basis function (RBF) can address the nonlinear relationships between the spectra and the reference data and is able to reduce the computational complexity of the training procedure. Therefore, RBF was employed as a kernel function, and consequently, the quality of the LS-SVM models depends on the hyperparameters gamma and sig2. The regularization parameter gamma determines the tradeoff between minimizing the training error and minimizing the model complexity. The parameter sig2 is the bandwidth of the RBF and implicitly defines the nonlinear mapping from the input space to some high dimensional feature space.

In recent years, many approaches for tuning the hyperparameters have been proposed in the literature, such as a gradient descent,[16] a grid search,[17] a covariance matrix adaptation evolution strategy (CMA-ES),[18] a genetic algorithm (GA),[19] and more recently, PSO. Developed by Kennedy and Eberhart in 1995,[20] PSO is a stochastic global optimization technique inspired by the social behavior of bird flocks; PSO possesses the capability to escape from local optima, is easy to implement, and has fewer adjustable parameters. In the PSO algorithm, a swarm consists of individuals, the so-called

particles, that represent a potential solution for the optimization problem. Each particle flies around a multidimensional search space with a velocity directing the flight and adjusts its position according to its own best previous experience (p_{best}) and the experience of all other members (g_{best}). All the particles have fitness values that are evaluated by predefined fitness functions. In this study, mean squared error (MSE) is set to be the fitness function, which varies with the LS-SVM parameters. The population is updated according to the fitness information, and the particles move towards areas with a better solution.

In this study, for the initial experiments, given a swarm of 100 particles, each particle $i(i \in \{1, 2, \ldots, 100\})$ is associated with a position vector $x_i = (x_1^i, \ldots, x_d^i, \ldots, x_D^i)$. D is the number of decision parameters in the optimization problem; in this study, $D = 2$ (gamma and sig2). The velocity of the ith particle is represented as $v_i = (v_1^i, \ldots, v_d^i, \ldots, v_D^i)$, and the ith particle accelerates toward the best fit locations according to the following equation:

$$v_d^i = \omega v_d^i + c_1 r_1 \times (p_{\text{best},d}^i - x_d^i) + c_2 r_2 \times (g_{\text{best},d}^i - x_d^i),$$

where x_d^i is the particle's coordinate in the Dth dimension, $p_{\text{best},d}^i$ is the location in the parameter space of the best fitness for a specific particle, $g_{\text{best},d}^i$ is the location in the parameter space of the best fitness for the entire swarm, c_1 and c_2 are learning factors and are chosen to be $c_1 = c_2 = 2$, r_1 and r_2 are random parameters between 0 and 1, and ω is the inertia weight, chosen here to be 0.5. After implementing PSO, the optimal hyperparameters gamma and sig2 for the LS-SVM model were obtained.

For SVM models, RBF was also employed as a kernel function. The regularization parameter gamma and the parameter sig2 were tuned with grid search. The LS-SVM and SVM algorithms were performed using the Matlab (version 7.5, The MathWorks, Inc., Natick, MA, USA).

PLS models with 1–10 principal components (PCs) were investigated. Leave-one-out cross-validation was employed, and the optimum number of PCs included in the model was determined by PRESS (predicted residual error sums of squares).[21] PLS computations were performed using TQ analyst software (version 8.0, Thermo Nicolet, USA).

Before developing BP-ANN models, the input dimensionality of the NIR spectra was reduced by principal component analysis (PCA). A two-layer BP network was employed, and several training of networks was performed with different numbers of input neurons (equal to the number of PCs) from 1 to 10 and hidden layer neurons from 1 to 30. The performance of each ANN model was evaluated by calculating the MSE value. Besides, the *tansig* and *purelin* functions were employed as transfer functions for BP networks; the *trainlm* algorithm was selected for training the ANN models.[4] The other training parameters were selected as default. Neural Network Toolbox for MATLAB was used for development of BP-ANN calibration models.

3. Results and Discussion

3.1. *HPLC analysis*

The dynamic courses of total flavonoids, puerarin, salvianolic acid B and icariin concentrations during the extraction processes are shown in Fig. 2. During the first extraction process, the time evolution curves of total flavonoids, puerarin and salvianolic acid B had a similar trend. It is easy to notice that the concentrations grew steadily and the rate of increase remained fairly constant. However, for icariin, the concentration increased quickly in the first 20 min. Then, in the remaining time of the first extraction process, the concentration grew slowly, and the rate of increase had slowed down noticeably.

Furthermore, compared with the salvianolic acid B, the concentrations of total flavonoids, puerarin and icariin in the final extracts was various among the six batches, revealing that there were some differences of the physicochemical parameters in different batches of the raw materials of Radix Puerariae (puerarin) and Folium Epimedii (icariin); by contrast, the quality of the raw material of Radix Salviae Miltiorrhizae (salvianolic acid B) was relatively stable.

3.2. *NIR spectral analysis*

To maximize the contributions of the investigated intermediate quality attributes to the quantitative models and simultaneously minimize the contribution of noises or other unwanted signals, it is necessary to identify the usable spectral regions prior to developing the calibration models. The NIR

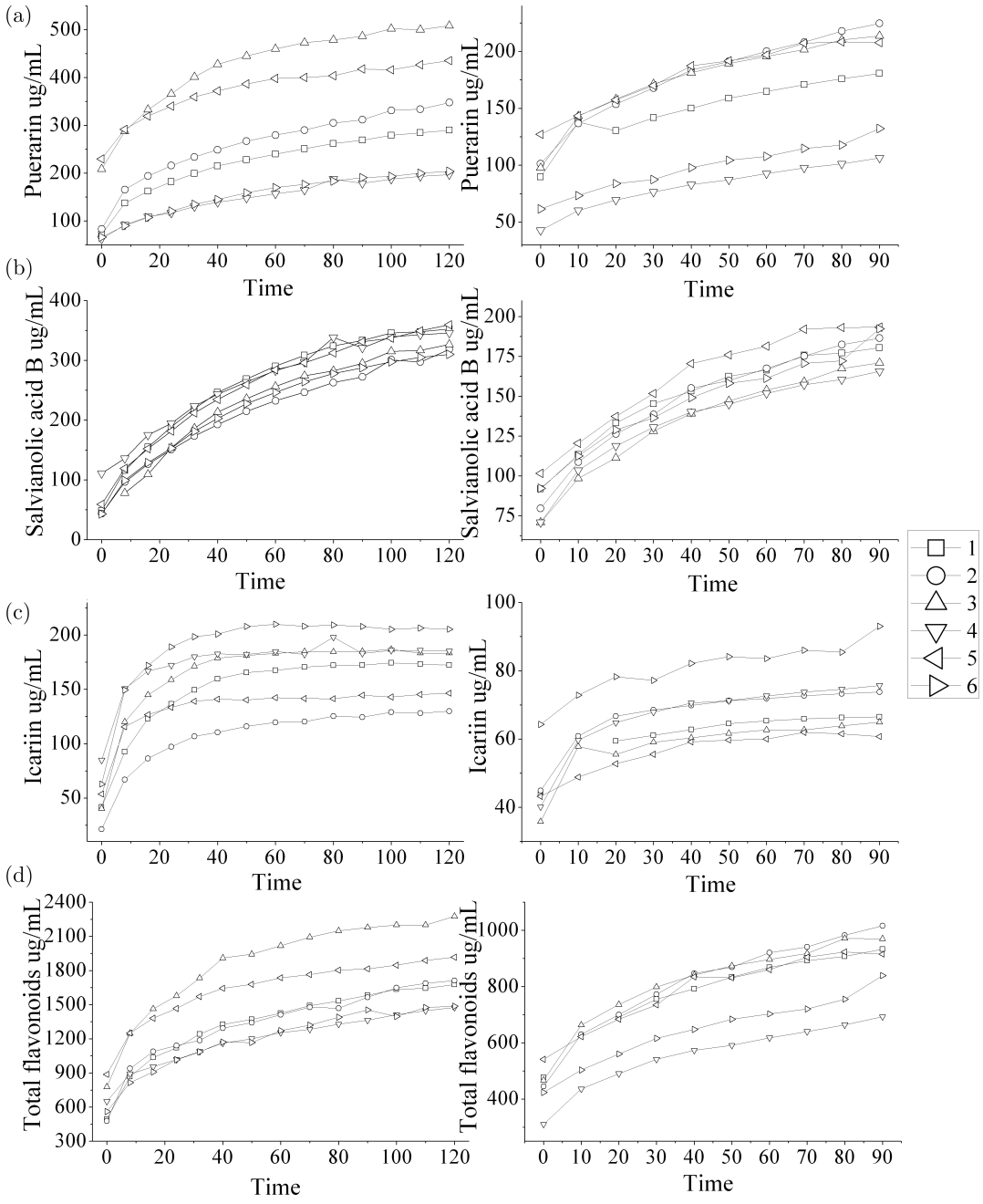

Fig. 2. Time evolution of puerarin (a), salvianolic acid B (b), icariin (c), and total flavonoids (d): 1 meant the first extraction process and 2 meant the second extraction process.

spectra exhibit intense absorption bands at 6944 cm^{-1} from the first O–H overtone and at 5155 cm^{-1} from the combination of stretching and deformation of the O–H group in water.[10] These two bands are typical of the NIR spectrum of an aqueous extraction solution and make it difficult to determine the active constituents with this technique because the bands mask any other bands present in these spectral ranges. Specific absorptions in the 4500–5000 cm^{-1} region correspond to combinations of fingerprint absorptions with the C–H, O–H and N–H stretching modes.[22] The 7500–12000 cm^{-1} region is assigned to the second and third overtones and is characterized by low intensity and low signal-to-noise ratio.[23] However, the 5400–6200 cm^{-1} region is useful, which encompasses bands originating from the first overtones of the C–H stretching modes.[24] The C–H

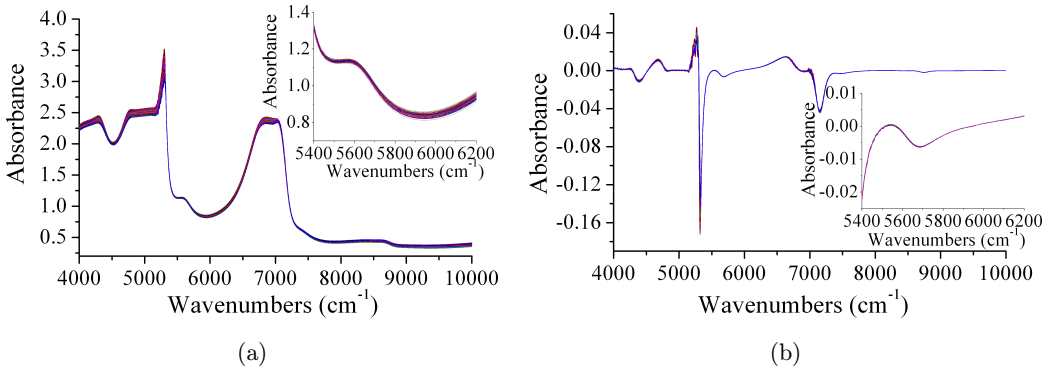

Fig. 3. Raw NIR spectra (a) and first derivative spectra smoothed by Savitzky–Golay filter (b).

stretching overtones in the 5400–6200 cm^{-1} region contribute substantially enough to form the basis for quantitative analysis of the four active components.[25,26] Therefore, for the total flavonoids, puerarin, salvianolic acid B and icariin models, the "fingerprint region" (4500–5000 cm^{-1}), the "water regions" (5000–5400 cm^{-1} and 6200–7500 cm^{-1})[27] and the region with low intensity and low signal-to-noise ratio (7500–10000 cm^{-1}) were excluded, and the spectral region actually employed was 5400–6200 cm^{-1} (see Fig. 3).

Data pretreatment was performed to reduce optical interference unrelated to the investigated quality attributes. Derivatives can reduce peak overlap and eliminate constant and linear baseline drifts.[28] However, for each derivative operation, the noise level increases slightly, the signal strength decreases dramatically and the signal-to-noise ratio decreases. Therefore, to eliminate the spectral differences from baseline shifts, first derivative spectra were adopted. To avoid enhancing the noise, all derivative spectra were smoothed using a Savitzky–Golay filter (cubic polynomial order and seven data points).

3.3. *Comparison of four regression methods*

After identifying the usable spectral regions, choosing appropriate preprocessing methods and detecting the outliers, the manipulated spectral data were correlated with the data measured by the reference assays. Using the optimum model parameters (number of PCs for PLS; number of input neuron and hidden layer neuron for BP-ANN; gamma and sig2 for LS-SVM and SVM), the quantitative calibration models for total flavonoids, puerarin, salvianolic acid B and icariin were established. The statistics of the developed models are listed in Table 1.

Based on the data in Table 1 (in bold), for the calibration and validation sets of the LS-SVM models, R^2 were all higher than 0.988, and the RMSEC (RMSEP) and RSEP were less than 0.042 and 8.778%, respectively. It is easy to see that, compared with the PLS, BP-ANN and SVM models, the calibration results of the LS-SVM models have much higher determination coefficients and lower RMSEC and RSEP values, indicating more satisfactory fitting results. And for the validation set, the LS-SVM models also provided the best prediction ability. This observation can be explained by the fact that the extraction solution of Yangxinshi tablet is complex and the data obtained by the NIR instrument contain both linear and nonlinear information. The LS-SVM method has the capability for linear and nonlinear multivariate calibration, and the PSO algorithm has been found to be fast and robust in solving nonlinear and nondifferentiable problems.[29] Therefore, the PSO-based LS-SVM method is more appropriate to handle the NIR spectral data of complex TCM solution in this study.

With respect to the RSEP values, BP-ANN had satisfactory calibration results, which were very close to the results of the PSO-based LS-SVM models; however, its prediction accuracy was inferior, especially the BP-ANN models of puerarin and salvianolic acid B. This observation can be explained by the fact that the ANN method has the tendency to overfit, and this behavior can lower the generalization ability of the network. However, PSO is a population-based stochastic approach that can avoid being trapped in a local optimum and can often find a global optimum solution.[30]

Table 1. Statistics of the optimal models using SVM, LS-SVM, PLS and BP-ANN methods.

		Calibration set				Validation set		
		Optimum model parameter	R^2	RMSEC	RSEP%	R^2	RMSEP	RSEP%
Total flavonoids	SVM	gamma $= 1.09 \times 10^6$; sig2 $= 2.02 \times 10^{-2}$	0.9703	0.078	4.940	0.9571	0.098	4.612
	LS-SVM	gamma $= 1.03 \times 10^7$; sig2 $= 3.39 \times 10^{-2}$	**0.9950**	**0.021**	**1.706**	**0.9894**	**0.042**	**2.998**
	PLS	PCs = 6	0.9766	0.070	5.852	0.9964	0.062	4.533
	BP-ANN	[8:10:1][a]	0.9926	0.026	2.158	0.9800	0.077	5.686
Puerarin	SVM	gamma $= 8.47 \times 10^5$; sig2 $= 1.98 \times 10^{-2}$	0.9111	0.031	12.617	0.8912	0.062	19.808
	LS-SVM	gamma $= 7.79 \times 10^6$; sig2 $= 3.40 \times 10^{-1}$	**0.9881**	**0.015**	**5.772**	**0.9778**	**0.018**	**5.820**
	PLS	PCs = 8	0.9545	0.023	10.286	0.9724	0.039	12.448
	BP-ANN	[8:10:1]	0.9908	0.010	5.157	0.7600	0.092	29.781
Salvianolic acid B	SVM	gamma $= 9.65 \times 10^5$; sig2 $= 1.56 \times 10^{-2}$	0.9078	0.025	11.710	0.8797	0.030	13.186
	LS-SVM	gamma $= 1.03 \times 10^7$; sig2 $= 3.39 \times 10^{-2}$	**0.9949**	**0.009**	**3.750**	**0.9803**	**0.012**	**5.646**
	PLS	PCs = 7	0.9734	0.013	6.208	0.9518	0.019	8.139
	BP-ANN	[9:10:1]	0.9905	0.008	3.387	0.6154	0.068	29.601
Icariin	SVM	gamma $= 1.01 \times 10^6$; sig2 $= 2.44 \times 10^{-2}$	0.9160	0.019	14.309	0.9249	0.020	17.886
	LS-SVM	gamma $= 3.47 \times 10^6$; sig2 $= 2.9 \times 10^{-3}$	**0.9886**	**0.007**	**4.498**	**0.9558**	**0.010**	**8.778**
	PLS	PCs = 7	0.9623	0.011	8.096	0.9656	0.040	36.592
	BP-ANN	[7:10:1]	0.9777	0.008	5.413	0.9887	0.019	17.813

a: [8:10:1] is the optimal topology for the BP-ANN model of total flavonoids, indicating 8 input neurons, 10 hidden layer neurons, and one output neuron.

Moreover, the total flavonoids models obviously presented the best performance and yielded the minimum prediction error. This phenomenon was mainly due to the high concentration of total flavonoids, ranging from 0.31 to 2.28 mg/mL in the calibration set; whereas, the concentrations of puerarin, salvianolic acid B and icariin were close to the accepted detection limit of NIR spectroscopy for natural products (0.1–0.01%).[31]

4. Conclusion

In this work, PSO-based LS-SVM, PLS, BP-ANN and SVM methods were investigated to build quantitative models for total flavonoids, puerarin, salvianolic acid B and icariin. Comparing the results, the PSO-based LS-SVM method provided much better fitting results and prediction accuracy and seems to be an interesting chemometric tool for quantitative prediction of extraction solution of Yangxinshi tablet.

References

1. D. Yan, Y. L. Qian, S. L. Tang, "Therapeutic effect of yangxinshi tablet in treating arrhythmia of patients with coronary heart disease with qi deficiency and blood stasis," *J. Nanjing TCM Univ.* **22**, 323–325 (2006).
2. Y. J. Wu, Y. Jin, H. Y. Ding, L. J. Luan, Y. Chen, X. S. Liu, "In-line monitoring of extraction process of scutellarein from Erigeron breviscapus (vant.) Hand-Mazz based on qualitative and quantitative uses of Near-infrared spectroscopy," *Spectroc. Acta A* **79**, 934–939 (2011).
3. Y. J. Wu, Y. Jin, Y. R. Li, D. Sun, X. S. Liu, Y. Chen, "NIR spectroscopy as a process analytical technology (PAT) tool for on-line and real-time monitoring of an extraction process," *Vib. Spectrosc.* **58**, 109–118 (2012).
4. Y. Jin, X. S. Liu, L. J. Luan, G. M. Qing, Y. Zhong, Y. J. Wu, "Rapid and quantitative detection method for acteoside during chromatographic purification of adhesive rehmannia leaf extract using near-infrared spectroscopy and chemometrics," *J. Near Infrared Spectrosc.* **21**, 43–53 (2013).
5. L. M. Reid, C. P. O'Donnell, G. Downey, "Recent technological advances for the determination of food authenticity," *Trends Food Sci. Technol.* **17**, 344 (2006).
6. X. C. Guo, J. H. Yang, C. G. Wu, C. Y. Wang, Y. C. Liang, "A novel LS-SVMs hyper-parameter selection based on particle swarm optimization," *Neurocomputing* **71**, 3211–3215 (2008).
7. Y. Jin, Z. Z. Wu, X. S. Liu, Y. J. Wu, "Near infrared spectroscopy as a process analytical technology (PAT) tool for on-line quantitative monitoring of

alcohol precipitation in combination with chemometrics," *J. Pharm. Biomed. Anal.* **77**, 32–39 (2013).
8. Y. Jin, Y. Kai, Y. J. Wu, X. S. Liu, Y. Chen, "Application of particle swarm optimization based least square support vector machine in quantitative analysis of extraction solution of safflower using near-infrared spectroscopy," *Chin. J. Anal. Chem.* **40**, 925–931 (2012).
9. J. A. K Suykens, J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.* **9**, 293–300 (1999).
10. W. Z. Lu, *Modern Near Infrared Spectroscopy Analytical Technology*, 2nd Edition, China Petrochemical Press, Beijing (2006).
11. T. Li, C. M. Zheng, T. Zhang, J. S. Tao, "Study on determination of pueraria flavonoids in extracts of Pueraria lobata (Willd.) Ohwi," *Shizhen Med. Mat. Med. Res.* **21**, 1088–1090 (2010).
12. D. Xiang, J. Berry, S. Buntz, P. Gargiulo, J. Cheney, Y. Joshi, B. Wabuyele, H. Q. Wu, M. Hamed, A. S. Hussain, M. A. Khan, "Robust calibration design in the pharmaceutical quantitative measurements with near-infrared (NIR) spectroscopy: Avoiding the chemometric pitfalls," *J. Pharm. Sci.* **98**, 1155–1166 (2009).
13. M. Blanco, A. Peguero, "An expeditious method for determining particle size distribution by near infrared spectroscopy: Comparison of PLS2 and ANN models," *Talanta* **77**, 647–651 (2008).
14. C. Cortes, V. N. Vapnik, "Support-vector networks," *Mach. Learn.* **20**, 273–297 (1995).
15. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Disc.* **2**, 121–167 (1998).
16. N. Ayat, M. Cheriet, C. Suen, "Automatic model selection for the optimization of SVM kernels," *Pattern Recogn.* **38**, 1733–1745 (2005).
17. C. W. Hsu, C. C. Chang, C. J. Lin, A Practical Guide for Support Vector Classification, Technique Report, National Taiwan University, Taipei (2003).
18. F. Friedrichs, C. Igel, "Evolutionary tuning of multiple SVM parameters," *Proc. 12th European Symp. Artificial Neural Networks (ESANN)*, M. Verleysen, Ed., pp. 519–524, Bruges, Belgium (2004).
19. C. Chatelain, S. Adam, Y. Lecourtier, L. Heutte, T. Paquet, "Multi-objective optimization for SVM model selection," *Proc. Ninth Int. Conf. Document Analysis and Recognition*, pp. 427–431, IEEE Computer Society, Washington DC (2007).
20. J. Kennedy, R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Networks*, pp. 1942–1948, IEEE Service Center, Piscataway, NJ (1995).
21. C. O. Chan, C. C. Chu, D. K. W. Mok, F. T. Chau, "Analysis of berberine and total alkaloid content in Cortex Phellodendri by near infrared spectroscopy (NIRS) compared with high-performance liquid chromatography coupled with ultravisible spectrometric detection," *Anal. Chim. Acta* **592**, 121–131 (2007).
22. K. Z. Liu, M. Shi, A. Man, T. C. Dembinski, R. A. Shaw, "Quantitative determination of serum LDL cholesterol by near-infrared spectroscopy," *Vib. Spectrosc.* **38**, 203–208 (2005).
23. N. Gierlinger, M. Schwanninger, R. Wimmer, "Characteristics and classification of Fourier-transform near infrared spectra of the heartwood of different larch species (Larix sp.)," *J. Near Infrared Spectrosc.* **12**, 113–119 (2004).
24. J. J. Kelly, J. B. Callis, "Nondestructive analytical procedure for simultaneous estimation of the major classes of hydrocarbon constituents of finished gasolines," *Anal. Chem.* **62**, 1444–1451 (1990).
25. F. A. Inon, R. Llario, S. Garrigues, M. de la Guardia, "Development of a PLS based method for determination of the quality of beers by use of NIR: Spectral ranges and sample-introduction considerations," *Anal. Bioanal. Chem.* **382**, 1549–1561 (2005).
26. J. W. Hall, A. Pollard, "Nearinfrared spectrophotometry–a new dimension in clinical-chemistry," *Clin. Chem.* **38**, 1623–1631 (1992).
27. M. Streefland, P. F. G. Van Herpen, B. Van de Waterbeemd, L. A. Van der Pol, E. C. Beuvery, J. Tramper, D. E. Martens, M. Toft, "A practical approach for exploration and modeling of the design space of a bacterial vaccine cultivation process," *Biotechnol. Bioeng.* **104**, 492–504 (2009).
28. W. J. McCarthy, *TQ Analyst User's Guide*, Thermo Nicolet Corp, Madison WI (2003).
29. H. W. Ge, Y. C. Liang, Y. Zhou, X. C. Guo, "A particle swarm optimization-based algorithm for job-shop scheduling problems," *Int. J. Comput. Methods* **3**, 419–430 (2005).
30. N. M. A. Latiff, C. C. Tsimenidis, B. S. Sharif, "Performance comparison of optimization algorithms for clustering in wireless sensor networks," in *Proc. IEEE Int. Conf. Mobile Adhoc and Sensor Systems*, Pisa, pp. 1–4 (2007).
31. E. Stark, K. Luchter, M. Margoshes, "Near-infrared analysis (NIRA)–a technology for quantitative and qualitative-analysis," *Appl. Spectrosc. Rev.* **22**, 335–399 (1986).