

Article

Application of Deep-Learning Algorithm Driven Intelligent Raman Spectroscopy Methodology to Quality Control in the Manufacturing Process of Guanxinning Tablets

Yi Tao ^{1,*}, Jiaqi Bao ¹, Qing Liu ¹, Li Liu ² and Jieqiang Zhu ^{1,*}¹ College of Pharmaceutical Science, Zhejiang University of Technology, Hangzhou 310014, China² Chiatai Qingchunbao Pharmaceutical Co., Ltd., Hangzhou 310023, China

* Correspondence: taoyi1985@zjut.edu.cn (Y.T.); zhujieqiang@zjut.edu.cn (J.Z.)

Abstract: Coupled with the convolutional neural network (CNN), an intelligent Raman spectroscopy methodology for rapid quantitative analysis of four pharmacodynamic substances and soluble solid in the manufacture process of Guanxinning tablets was established. Raman spectra of 330 real samples were collected by a portable Raman spectrometer. The contents of danshensu, ferulic acid, rosmarinic acid, and salviolic acid B were determined with high-performance liquid chromatography-diode array detection (HPLC-DAD), while the content of soluble solid was determined by using an oven-drying method. In the establishing of the CNN calibration model, the spectral characteristic bands were screened out by a competitive adaptive reweighted sampling (CARS) algorithm. The performance of the CNN model is evaluated by root mean square error of calibration (RMSEC), root mean square error of cross-validation (RMSECV), root mean square error of prediction (RMSEP), coefficient of determination of calibration (R_c^2), coefficient of determination of cross-validation (R_{cv}^2), and coefficient of determination of validation (R_p^2). The R_p^2 values for soluble solid, salviolic acid B, danshensu, ferulic acid, and rosmarinic acid are 0.9415, 0.9246, 0.8458, 0.8667, and 0.8491, respectively. The established model was used for the analysis of three batches of unknown samples from the manufacturing process of Guanxinning tablets. As the results show, Raman spectroscopy is faster and more convenient than that of conventional methods, which is helpful for the implementation of process analysis technology (PAT) in the manufacturing process of Guanxinning tablets.

Keywords: Guanxinning tablets; Raman spectroscopy; convolutional neural network; process analytical chemistry; extraction process monitoring; intelligent process analysis



Citation: Tao, Y.; Bao, J.; Liu, Q.; Liu, L.; Zhu, J. Application of Deep-Learning Algorithm Driven Intelligent Raman Spectroscopy Methodology to Quality Control in the Manufacturing Process of Guanxinning Tablets. *Molecules* **2022**, *27*, 6969. <https://doi.org/10.3390/molecules27206969>

Academic Editors: Angelo Antonio D'Archivio and Alessandra Biancolillo

Received: 27 September 2022

Accepted: 12 October 2022

Published: 17 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Quality control in the manufacturing process is an important issue to guarantee the quality of end-products of botanical drugs [1,2]. In 2004, the U.S. Food and Drug Administration (FDA) issued the process analytical technology (PAT) industry guide, pointing out that a variety of methods exist as an overall quality-control system for PAT, which are used to monitor key quality attributes of raw materials and intermediates in real time, and to provide a reliable guarantee for the quality of end-products [3]. Given this guideline, in-process analytical methods and techniques are required to provide real-time quality information on botanical drugs.

The Guanxinning tablet is a botanical drug which is clinically used for the treatment of coronary heart disease and angina pectoris. The raw materials of the Guanxinning tablet are *Salvia miltiorrhiza* and *Ligusticum chuanxiong hort.* The extract of *Salvia miltiorrhiza* has the pharmacological effects of anti-platelet aggregation, anti-thrombosis, promoting fibrin degradation, anti-myocardial ischemia, and antioxidant [4,5], whereas the extract of *Ligusticum chuanxiong hort.* can dilate blood vessels, increase coronary flow, and improve microcirculation [6]. The conventional quality-control method used in the manufacture of Guanxinning tablets is high-performance liquid chromatography, which is used for

determining the bioactive compounds, such as danshensu, ferulic acid, rosmarinic acid, and salvianolic acid B. The disadvantage of the HPLC method is that it is time-consuming and consumes large amounts of organic solvents. Meanwhile, the HPLC method is unsuitable for real-time monitoring. To address this issue, Raman spectroscopy serves as a suitable alternative.

Compared with conventional PAT techniques, Raman spectroscopy has the merits of being non-destructive, fast, and portable. Also, Raman spectroscopy has been widely used in molecular fingerprinting [7,8], pathogenic bacteria discrimination [9], tumor diagnosis [10], and so on. The Raman spectrum of water is very weak, so, Raman spectroscopy is suitable for detection uses in liquid samples [11,12]; however, the application of Raman spectroscopy to quality control in the production process of botanical drugs is still rare.

Since Raman spectroscopy alone is still unable to achieve real-time detection, it needs to be combined with machine-learning algorithms [13]. The convolutional neural network algorithm (CNN), which is a deep-learning algorithm, has received increasing attention. As compared to the traditional machine-learning algorithms, such as partial least squares regression (PLSR) and support vector machine regression (SVR), the deep-learning algorithm can simplify the spectral deconvolution process and improve the model accuracy, making it very suitable for processing big data in the manufacturing process.

In this paper, an intelligent deep-learning algorithm driven Raman spectroscopy analysis methodology was established for the rapid and non-destructive determination of the contents of four bioactive compounds and soluble solid in water extract samples of Guanxinling tablets, which were collected from the manufacturing process. The model performance of CNN was compared with that of PLSR and SVR.

2. Results and Discussion

2.1. Determination of Bioactive Ingredients by HPLC-DAD

A reliable HPLC-DAD method was established to determine the four bioactive ingredients. The chemical structures of the four analytes are shown in Figure 1. Danshensu, ferulic acid, rosmarinic acid, and salvianolic acid B have good UV-absorption characteristics at 288 nm. The representative chromatograms of real sample and mixed standard solutions are displayed in Figure 2. Peaks 1–4 represent danshensu, ferulic acid, rosmarinic acid, and salvianolic acid B.

335 real samples were analyzed by the HPLC-DAD method. All four bioactive compounds were baseline separated and could be accurately determined. Calibration curves, correlation coefficients, linearity ranges, and LOD and LOQ data are shown in Table S1. All four major compounds displayed good correlation coefficient values (r^2) in the range of 0.9995–0.9997. The LODs and LOQs of the four major compounds were in the range from 0.2042 to 0.5313 $\mu\text{g}/\text{mL}$ and from 0.6807 to 1.7709 $\mu\text{g}/\text{mL}$, respectively. The method was fully validated. The precision, repeatability, stability, and recovery of the method are shown in Table S2. The RSDs of intra-day and inter-day precisions of the method were determined to be in the range of 0.07–0.47% and 1.10–1.49%. The repeatability and stability of the method were determined as from 0.15% to 1.07% and from 0.11% to 1.03%. The overall recoveries ranged from 99.80% to 103.62%.

2.2. Determination of Soluble Solid by an Oven-Drying Method

The content of soluble solid is an important indicator of the water extracts. It is also necessary to establish a reliable reference method for determining soluble solid content. 335 samples were assayed by using the oven-drying method. The soluble solid of the water-extract samples ranged from 418.7 $\mu\text{g}/\text{mL}$ to 4882.7 $\mu\text{g}/\text{mL}$.

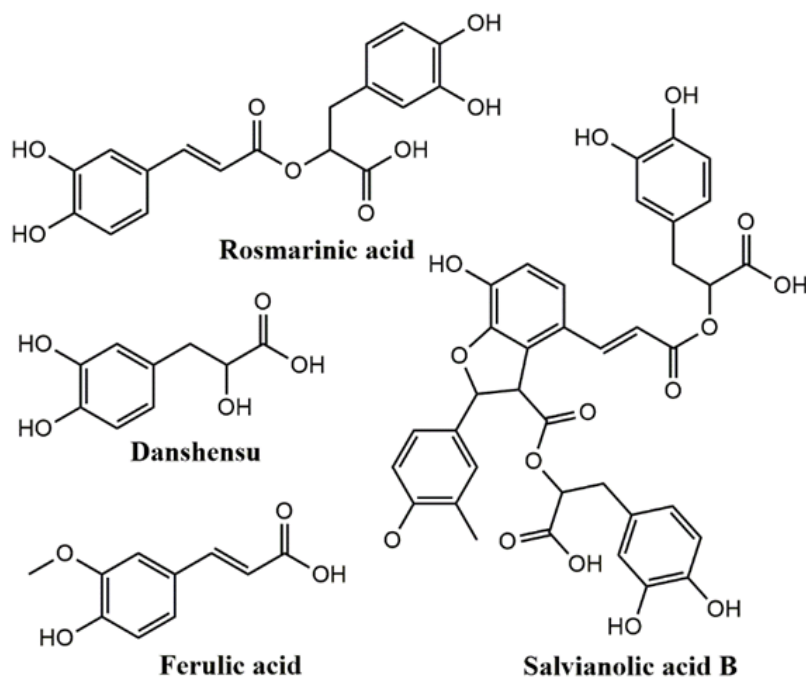


Figure 1. The chemical structures of the four investigated analytes.

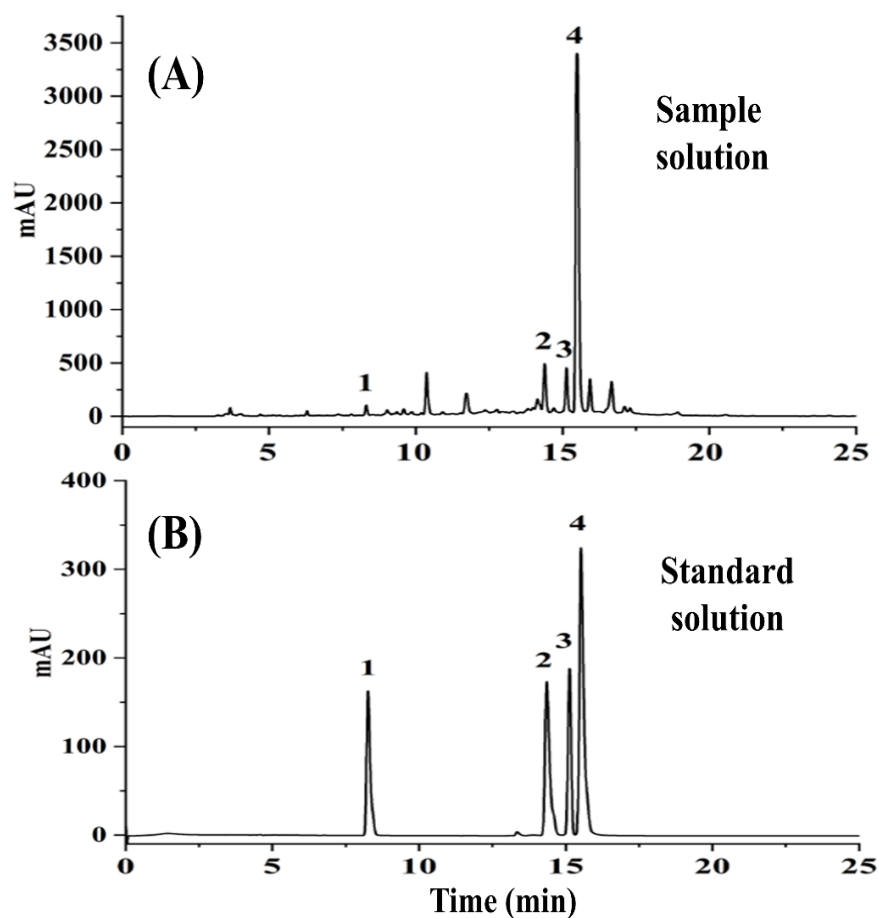


Figure 2. Representative HPLC chromatograms of sample solution (A) and standard solution (B).

2.3. Pretreatment of Raman Spectra

The raw Raman spectra of Guanxinning water extract are shown in Figure 3A. Savitzky–Golay (S–G) smoothing and Minmax linear regression were used (see Figure 3B,C). It can be seen in the raw spectrum that the major Raman peaks are 1000, 1250, and 1500 cm^{-1} . The major Raman peaks were assigned by comparing with the literature [14,15]. The Raman peak at 1000 cm^{-1} can be ascribed to Ar ring stretching. The Raman peak at 1250 cm^{-1} can be ascribed to asymmetric stretching of the C–O–C bond. The Raman peak at 1500 cm^{-1} can be ascribed to C=C bond stretching.

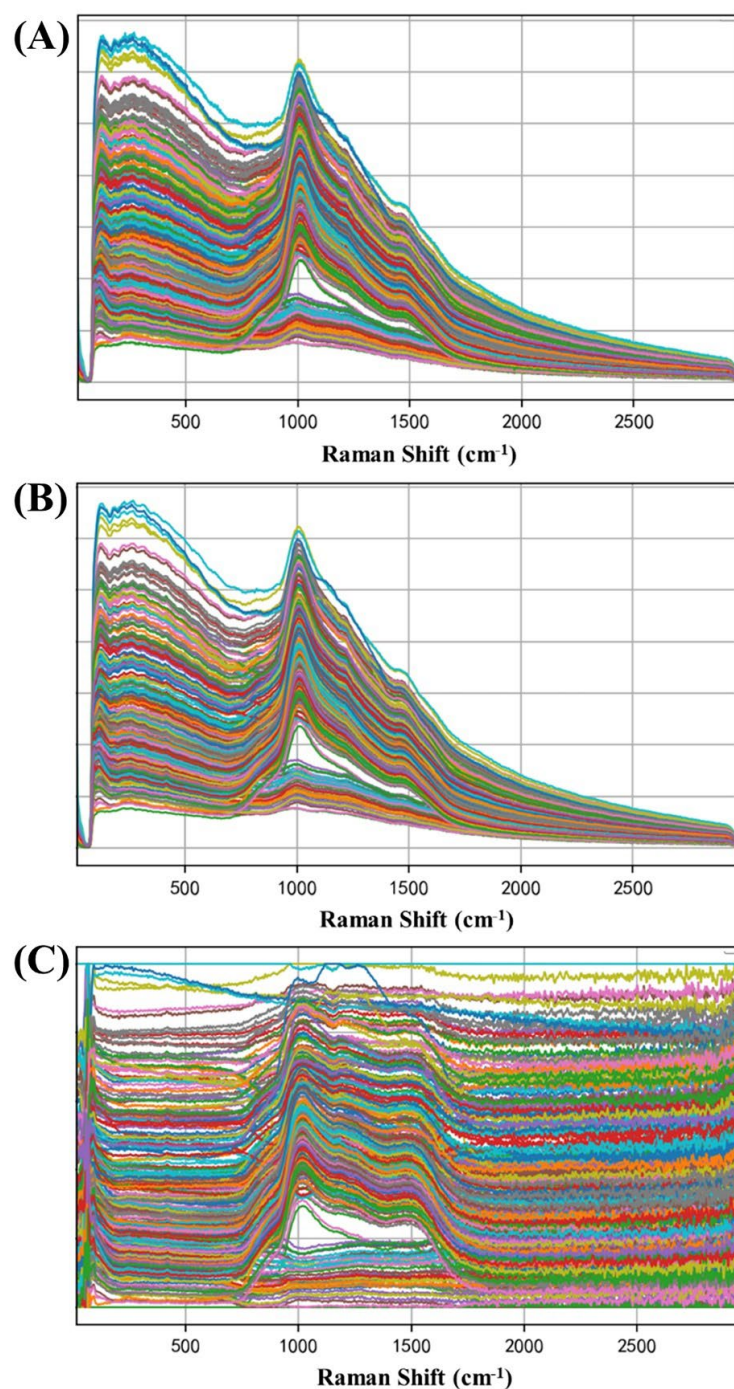


Figure 3. Raw Raman spectra (A) and spectra preprocessed by S–G (B) and Minmax (C) of water extract samples.

2.4. Removal of Abnormal Spectra

The number of abnormal Raman spectra in 335 water-extract samples of Guanxinning was removed by using the Mahalanobis distance method. The Mahalanobis distance distribution of the samples is shown in Figure 4. Five abnormal samples (No. 4-1-5, No. 4-3-1, No. 6-3-9, No. 7-2-3, and No. 7-2-8) were identified and removed; therefore, the number of water-extract samples of Guanxinning for building a quantitative calibration model was 330.

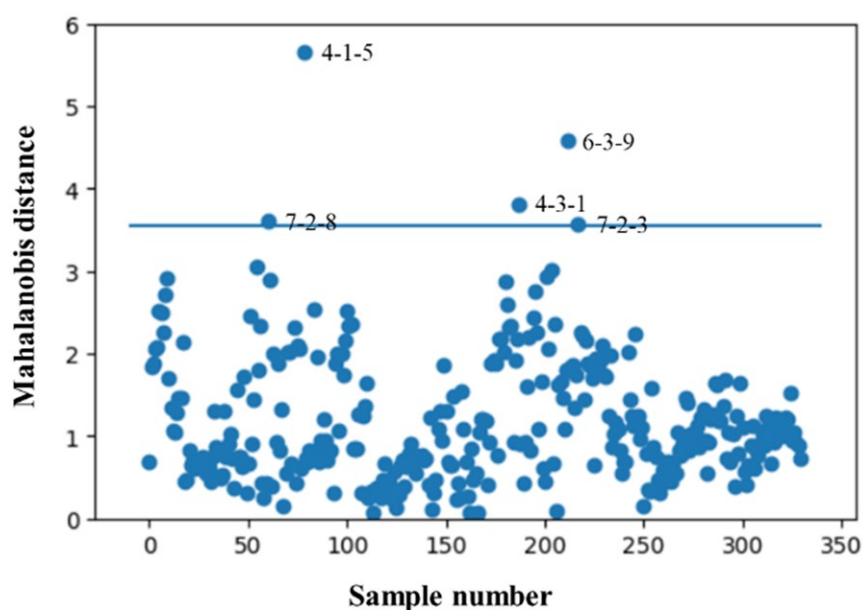


Figure 4. Mahalanobis distance distribution diagram of the samples.

2.5. Determination of Variable Selection Methods

Using the Kennard–Stone (K–S) algorithm, 330 real samples were divided into calibration set and validation set by 4:1. For each model, the calibration set consists of 264 samples, and the remaining 66 samples belong to the validation set. Table S3 lists the statistical values of the content of the four bioactive compounds in the calibration set and validation set. The calibration set covers a large range, which helps to build a stable and robust calibration model.

Four variable selection methods were used to select the characteristic bands of Raman spectra. CNN, PLSR, and SVR models were established, respectively. RMSEC, RMSEP, R_c^2 , and R_p^2 were used to evaluate the performance of these models. Taking the PLSR model, for example, the calculation results are shown in Table S4. Among the four feature band selection algorithms, CARS shows the best prediction performance for danshensu, ferulic acid, rosmarinic acid, salvianolic acid B, and soluble solid with an R_p^2 at 0.6382, 0.8483, 0.9457, 0.8696, and 0.9282; thus, the CARS algorithm is adopted as the Raman spectral feature band selection method.

In the CARS algorithm, the Monte Carlo sampling rate was set to 0.8, and the sampling number was 50. Figure 5 represents a process diagram of CARS to extract variables. Taking rosmarinic acid as an example, when the number of times increases from 0 to 50, both the number of extractions changes (see Figure 5A) and the RMSECV values (see Figure 5B) change, but the change trends are obviously different. When the number of samples increases from 0 to 10, the number of selected variables decreases rapidly, which is the fast screening stage, and this process removes a lot of invalid information. When the number of samples is greater than 10, the number of variables shows a slow downward trend, which is the fine screening stage. When the number of samples is 16, the RMSECV value is the smallest, and the number of variables at this time is the optimal variable set.

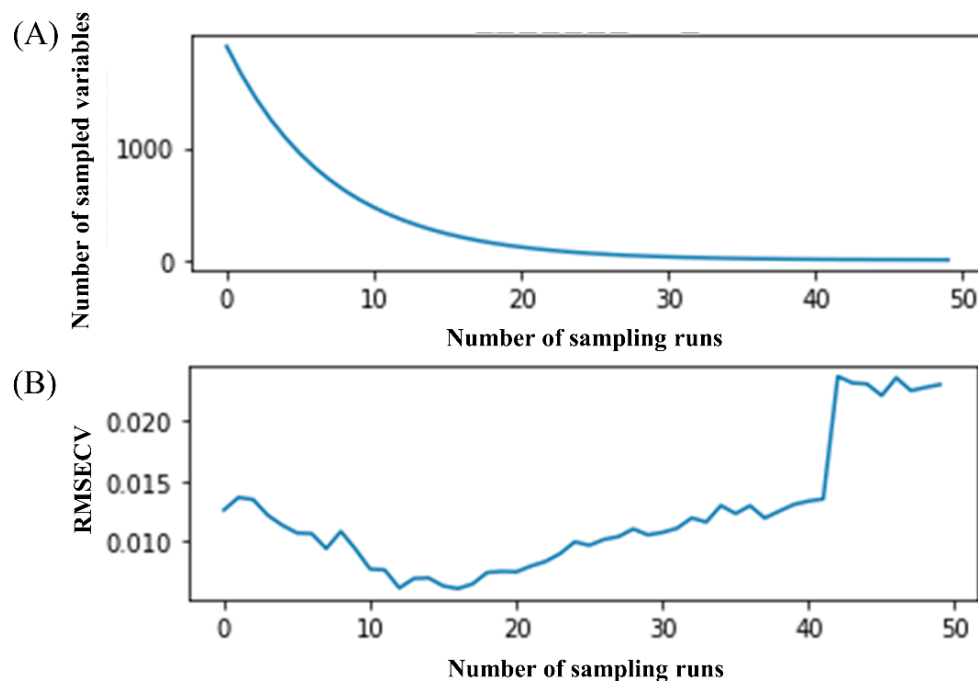


Figure 5. The process of selecting characteristic bands of Raman spectrum by CARS. (A) Number of sampled variables changed according to the number of sampling runs; (B) RMSECV values changed according to the number of sampling runs.

2.6. Comparison of Different Calibration Models

The performance parameters of the different calibration models established with the optimal band selecting method are listed in Table 1. According to the performance parameters, CNN, PLSR, and SVR were compared. It was worth mentioning that PLSR and SVR algorithms required preprocessing of the data. The calibration model of SPA-SVR showed the worst predictive ability, with an R_p^2 of -1.4600 and -0.0021 . The calibration models of CARS-PLSR and CNN had modest predictive ability, with an R_p^2 of 0.6382 , 0.8483 , 0.9457 , 0.8696 , and 0.9282 , and 0.8212 , 0.7163 , 0.8450 , 0.8544 , and 0.8554 . The calibration model of CARS-CNN had the best predictive ability with an R_p^2 of 0.8458 , 0.8667 , 0.8491 , 0.9246 , and 0.9415 . The RMSEP of the calibration model of CARS-CNN was the lowest. Correlation diagrams of predicted values and measured values of the five attributes of samples are shown in Figure 6.

2.7. Application to Three Batches of Unknown Samples

The established method is used for routine analysis in the production process, with which three batches of unknown samples of Guanxinning water extract were acquired by a portable Raman spectrometer. The Raman spectra of the unknown samples were corrected and inputted into the established CARS-CNN model. The contents of the four bioactive compounds and soluble solid were obtained at the same time. The three different batches of samples were analyzed with this model. Figure 7 shows the applications of Raman spectroscopy and CARS-CNN model to the unknown samples. The content of the main compounds is monitored and controlled during the production process through this method, allowing us to check whether the end-product meets the required standard and it further ensures the quality of the end-product. From Figure 7, we agreed that feasibility and superiority of CARS-CNN for the prediction of unknown samples is not apparent. The performance of deep learning highly depends on the size of the samples. The larger the sample size, the better the performance of the model will be. In our work, the sample size was 330, which is still too small for the CARS-CNN model. We believe the incorporation of

more data of the samples into the model will unambiguously improve the performance and show the superiority of deep learning.

Table 1. Comparisons between performance parameters of four quantitative calibration models established with different regression algorithms.

Algorithms	Objectives	Calibration		Cross-Validation		Prediction	
		R_c^2	RMSEC	R_{cv}^2	RMSECV	R_p^2	RMSEP
SPA-SVR	Danshensu	−1.3462	0.0205	−1.6672	0.0176	−1.4600	0.0265
	Ferulic acid	−4.6801	0.0145	−5.2301	0.0423	−0.0021	0.0158
	Rosmarinic acid	0.5132	0.0424	0.4102	0.0644	0.3652	0.0448
	Salvianolic acid B	0.8718	0.2929	0.7189	0.2034	0.8100	0.3580
	Soluble solid	0.8185	0.6091	0.7033	0.7651	0.7189	0.7647
CARS-PLSR	Danshensu	0.9979	0.0011	0.9223	0.0031	0.6382	0.0163
	Ferulic acid	0.9843	0.0045	0.9456	0.0025	0.8483	0.0057
	Rosmarinic acid	0.9904	0.1539	0.9754	0.1634	0.9457	0.0092
	Salvianolic acid B	0.9958	0.0584	0.9192	0.1345	0.8696	0.3749
	Soluble solid	0.9904	0.1539	0.9312	0.2745	0.9282	0.4512
CNN	Danshensu	0.8694	0.0086	0.8423	0.0035	0.8212	0.0096
	Ferulic acid	0.8893	0.1197	0.7356	0.0768	0.7163	0.0213
	Rosmarinic acid	0.9285	0.0098	0.8749	0.0167	0.8450	0.0143
	Salvianolic acid B	0.8906	0.2988	0.8876	0.3758	0.8544	0.3371
	Soluble solid	0.8857	0.5389	0.8831	0.5775	0.8554	0.5812
CARS-CNN	Danshensu	0.9893	0.0024	0.9423	0.0051	0.8458	0.0094
	Ferulic acid	0.9875	0.0016	0.9335	0.0035	0.8667	0.0084
	Rosmarinic acid	0.9551	0.0078	0.9464	0.0079	0.8491	0.0145
	Salvianolic acid B	0.9943	0.1953	0.9213	0.1886	0.9246	0.2528
	Soluble solid	0.9526	0.1197	0.9384	0.3188	0.9415	0.3861

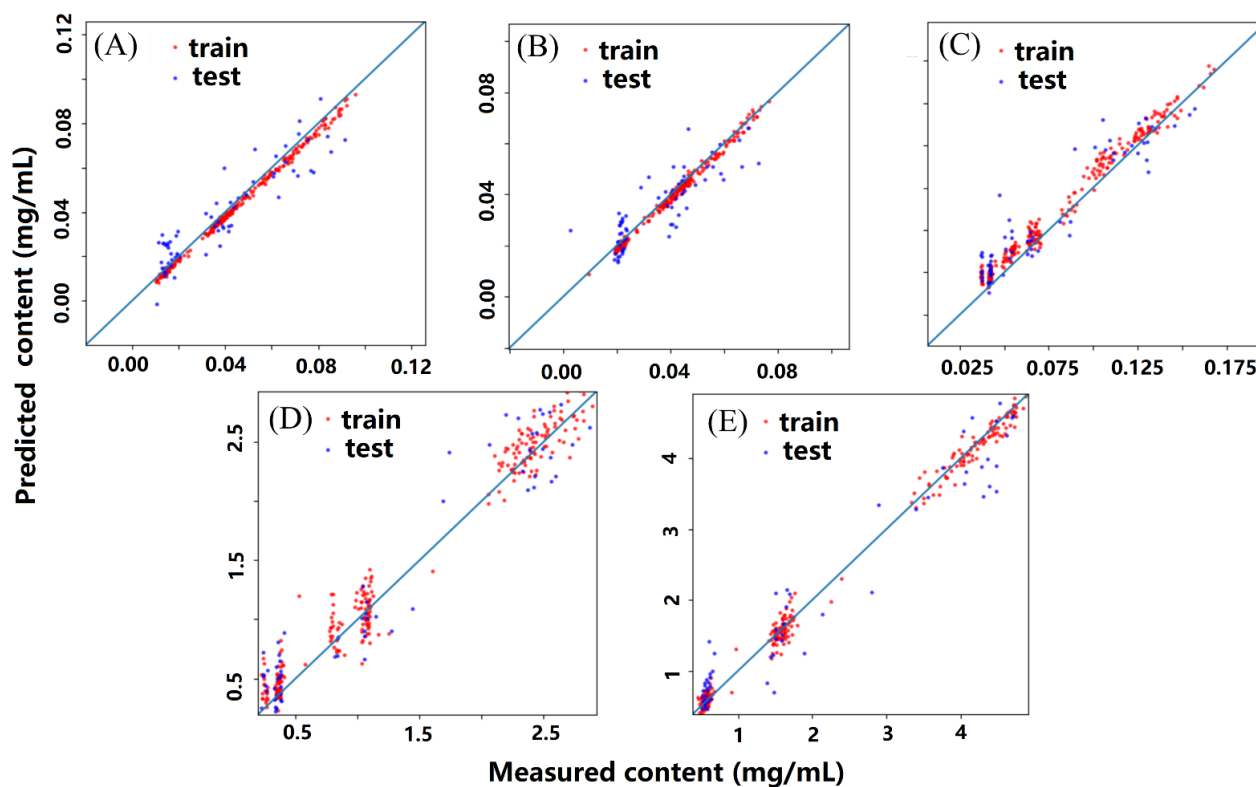


Figure 6. Correlation diagram of predicted values and measured values of the five attributes: (A) danshensu, (B) ferulic acid, (C) rosmarinic acid, (D) salvianolic acid B, and (E) soluble solid.

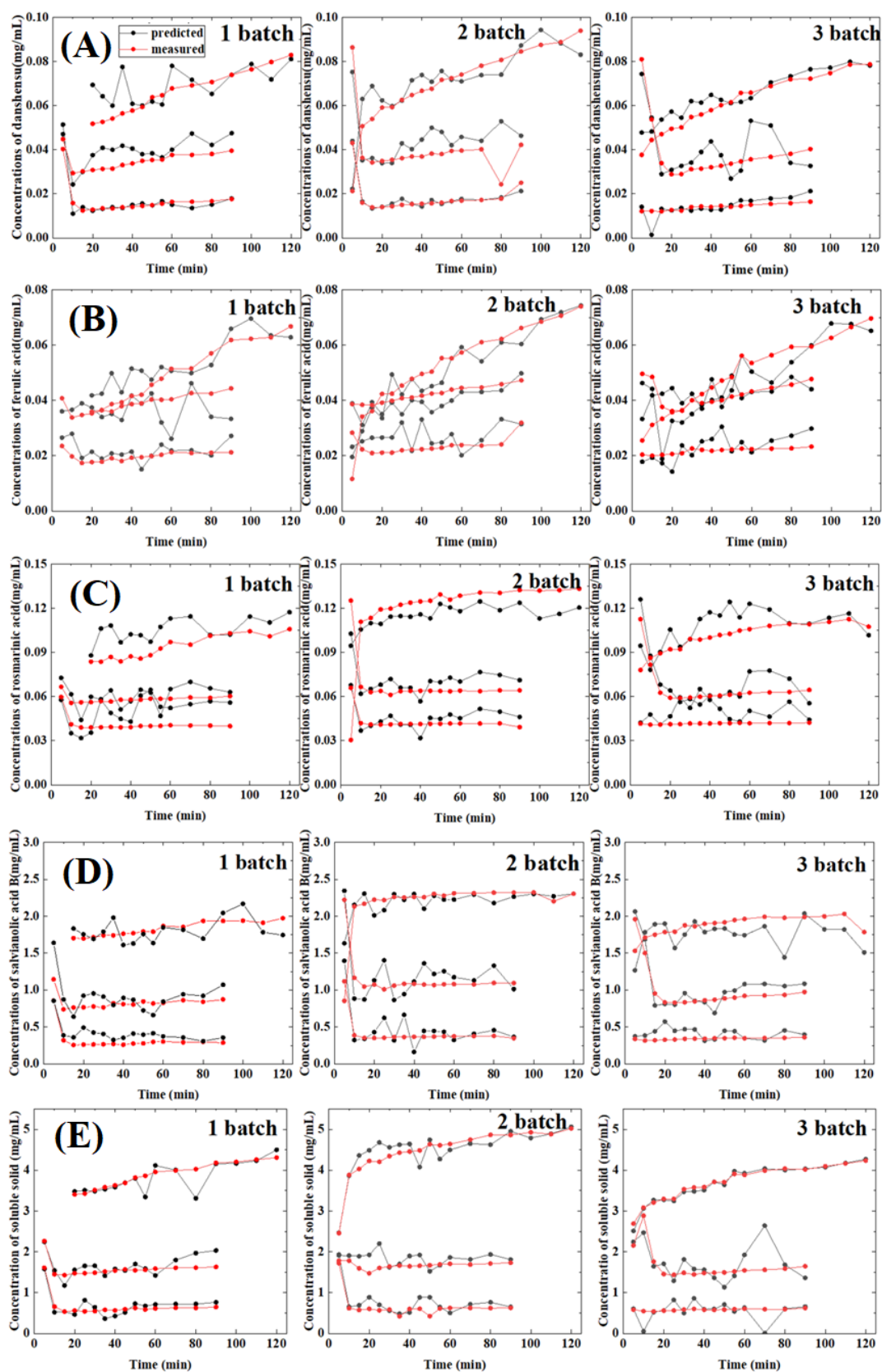


Figure 7. Contents of the five key attributes of three batches of unknown samples predicted with the established CARS-CNN model: (A) danshensu, (B) ferulic acid, (C) rosmarinic acid, (D) salvianolic acid B, and (E) soluble solid.

3. Materials and Methods

3.1. Sample Collection

Water-extract samples of Guanxinning tablets were collected from a Chinese medicine pharmaceutical factory (Zhengda Qingchunbao Pharmaceutical Co., Zhejiang, China) in Deqing. In the extraction process during the production of the Guanxinning tablets, reflux extraction was carried out three times. The extraction time of the first reflux process was 2 h. The extraction time of the second reflux process was 1.5 h. The extraction time of the third reflux process was 1.5 h. 10 mL samples were collected every 5 min for the first 1 h, and 10 mL samples were collected every 10 min for the next 1 h. When the crude drug enters the second and third reflux processes, 10 mL samples were collected every 5 min for the first 1 h, and 10 mL samples were collected every 10 min for the next 0.5 h. A total of seven batches of samples (335 samples) were collected from the Guanxinning water-extraction module. The port of extractor was spun off at certain time points and the water extract was poured into a small beaker. Then, the water extract was transferred to a centrifuge tube. The flow chart of this study is shown in Figure 8.

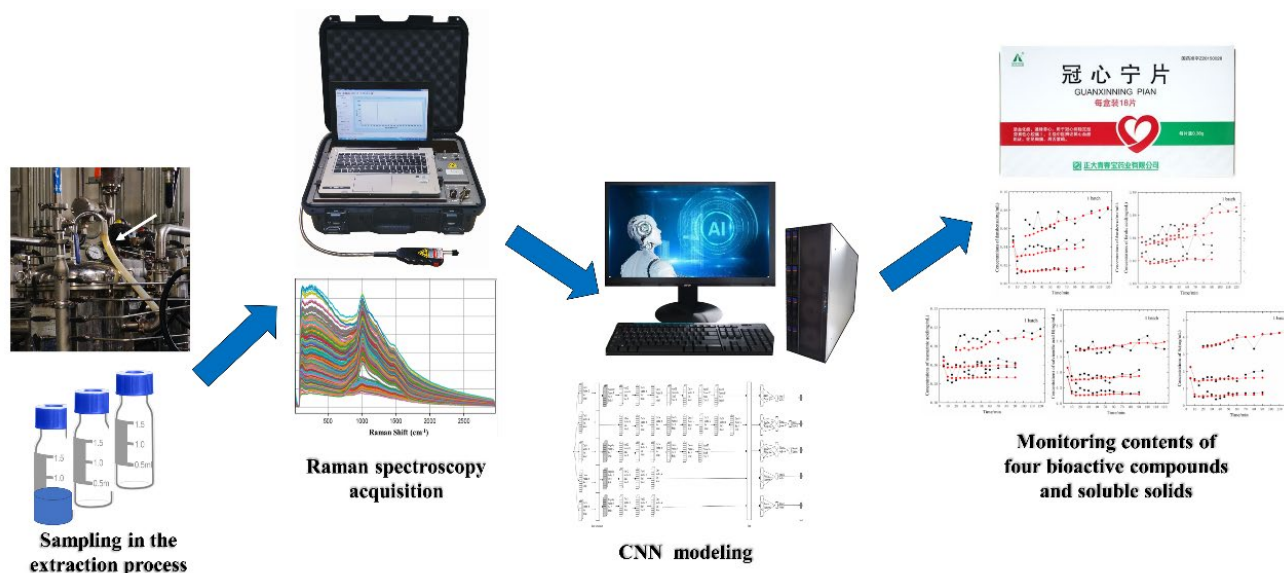


Figure 8. Schematic representation of the convolutional neural network enabled intelligent Raman spectroscopy methodology for monitoring the contents of four bioactive compounds and soluble solids in the water extracts of Guanxinning tablets.

3.2. HPLC-DAD Analysis

In order to determine the concentration of four bioactive compounds in Guanxinning water extract, a high-performance liquid chromatography method was established. The water extract was centrifuged at 13,000 rpm for 10 min. Then, the supernatant was sent for analysis under the following chromatographic conditions. An Agilent 1260 high-performance liquid chromatography system (Agilent Technologies, Santa Clara, CA, USA) was used, including a quaternary pump, a sample vial injector, a column oven, and a diode array detector (DAD). The column was Hanbon Sci & Tech Hedera ODS-2 (4.6 × 250 mm, 5 μm), and the mobile phases consisted of (A) 0.1% HCOOH-H₂O (*v/v*) and (B) acetonitrile. The gradient elution procedure was as follows: initial 95% (A); 0–12 min, 5–38% (B); 12–20 min, 38–48% (B); 20–35 min, 48–100% (B). The re-equilibration duration between single runs was 6 min. The column temperature was 36 °C and the flow rate was 0.8 mL/min. The detection wavelength of danshensu, ferulic acid, rosmarinic acid, and salvianolic acid B was 288 nm. LODs and LOQs were determined by using diluted standard solution when the signal-to-noise ratios (S/N) of the standard substances were about 3 and 10, respectively. Variations were expressed by relative standard deviations (RSD).

3.3. Oven-Drying Method

In order to determine the content of soluble solid in the Guanxinning water extract, an oven-drying method was adopted. Guanxinning water extract was centrifuged at 2500 rpm for 10 min. Then, about 3 mL of the supernatant was placed in a flat weighing bottle, evaporated to dryness in a water bath, and then placed in a 105 °C oven for 6 h. Finally, the bottle was taken out, placed in a desiccator to cool for 1 h, and weighed. The soluble solid content was calculated according to Formula (1), where S_c is the soluble solid content of the extract, W is the quality of the extract, W_2 is the total mass of the sample and weighing bottle after drying, and W_1 is the mass of the weighing bottle.

$$S_c = \frac{W_2 - W_1}{W} \times 100\% \quad (1)$$

3.4. Raman Spectra Acquisition

The Raman spectrum was collected by a Rapid OLRaman-2 portable Raman spectrometer equipped with a Raman fiber probe, a CCD detector, and a laser emitter (power 400 mW, wavelength 785 nm). It is a dispersive (with a grating) type of instrument. The acquisition parameters are as follows: wavenumber range 176–3500 cm^{-1} , resolution 2.83 cm^{-1} , acquisition time 500 ms, and samples were collected three times each. The Raman spectrometer was controlled by a compatible flat panel, and the “Pharmaceutical” software (Version 1.0) was used for data acquisition.

3.5. Removal of Abnormal Samples

In addition to the sample information, the data collected by the Raman spectrometer also include abnormal spectra that may have been generated due to errors of instrument, method, environment, or manual operation during the collection process. In order to obtain a reliable, accurate, and stable quantitative model, it is necessary to identify and remove abnormal spectra before modeling. In order to eliminate the interference of abnormal spectra, Mahalanobis distance method was used.

3.6. Feature Band Filtering

The Raman signal shift of the Raman spectrometer is 176–3500 cm^{-1} . In order to best utilize the effective spectrum, the optimal characteristic bands should be screened out during the calibration process. To screen out the optimal spectral bands, competitive adaptive reweighted sampling (CARS), Uninformative Variable Elimination (UVE), Successive Projections Algorithm (SPA), and Synergy Interval Partial Least Square (siPLS) toolbox were used [16]. The performances of different screening algorithms were compared, and the best feature band screening algorithm was selected.

3.7. Determination of Variable Selection Methods

The algorithms used for building the calibration models were PLSR, SVR, and CNN. The principle and application of these algorithms were well documented in the references [17–19]. The architecture of CNN model is shown in Figure 9.

The construction detail of the CNN model is as follows. First, a convolution layer was created. The parameters of the convolution layer were as follows: 32 filters, the filter window size was 3×3 , the scanning window moved with a step size of 1 each time, and the rectified linear units (ReLU) activation function was applied. Second, a batch-normalization layer was created. Third, a maximum pooling layer was created. The number of filters in the pooling layer was the same as that of the convolution layer 1. The filter window size was 2×2 , and the scanning window moved with a step size of 1 each time. There was no maxpooling layer for ferulic acid. Fourth, four convolutional layers with parameters setting to 16, 3, and 1 were created successively. The ReLU activation function was applied. After that, a convolution layer with parameters setting to 32, 3, and 1 was created and the ReLU activation function was applied. Then, a convolution layer with parameters setting to 64, 3,

and 1 was created and the ReLU activation function was applied. Finally, a flattened layer was created. Then, two fully-connected layers were created for danshensu, salvianolic acid B, and soluble solid, whereas three fully-connected layers were created for ferulic acid and rosmarinic acid. The number of output neurons is 1, and the linear activation function was applied. The mean squared error loss function was chosen and Adam ($\text{lr} = \text{le} \times -4$) was used as the optimizer. The specified batch size was 50. The number of iterations was 200. If the loss was not improved after 40 iterations, Keras would stop training.

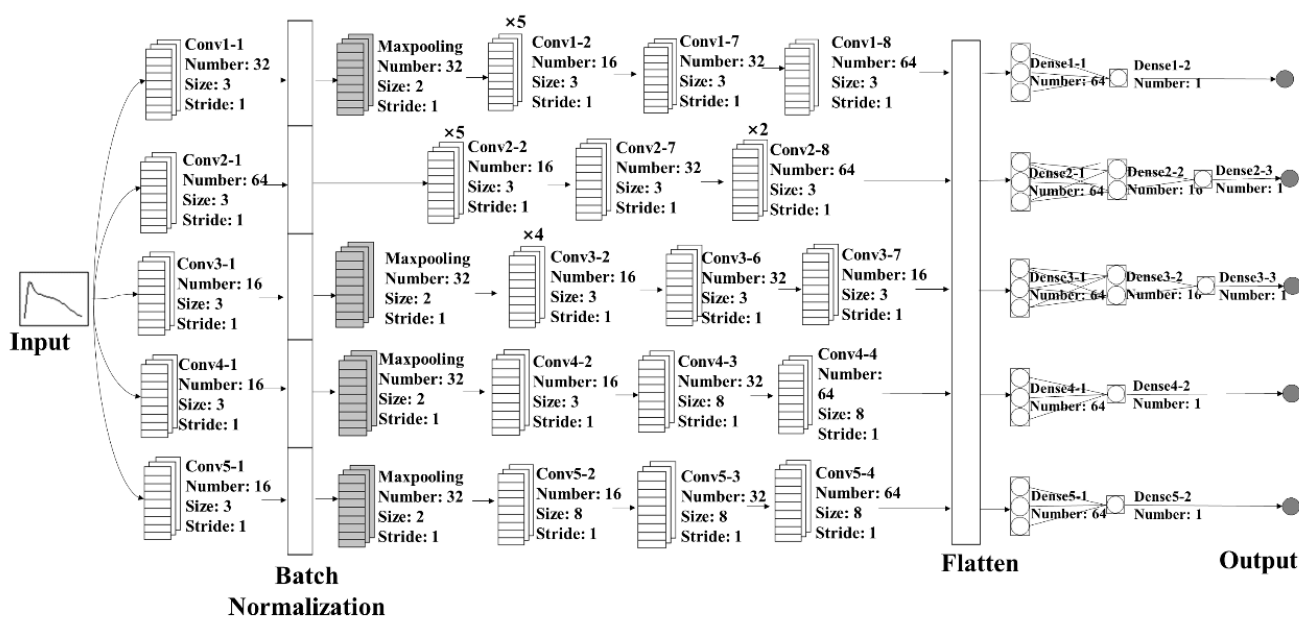


Figure 9. Architectures of the established CNN model.

The root mean square error of calibration (RMSEC), root mean square error of cross-validation (RMSECV), root mean square error of prediction (RMSEP), correlation coefficient of calibration (R_c^2), correlation coefficient of cross-validation (R_{cv}^2), and correlation coefficient of validation (R_p^2) were used to evaluate the performances of the above models. The detailed calculation formulas of the above parameters can be found in the literature [20].

4. Conclusions

An intelligent Raman spectroscopy methodology for the simultaneous determination of danshensu, ferulic acid, rosmarinic acid, salvianolic acid B, and soluble solid in water extract of Guanxinning was established. The calibration model has been validated with satisfactory R_p^2 values. The method has been successfully applied to the monitoring of the contents of pharmacodynamic substances and soluble solid in water extracts of Guanxinning tablets, which improves the efficiency of quality control, and may replace the cumbersome reference method. The model needs to be updated to ensure robustness for long-term use in industrial manufacturing. To the best of our knowledge, this study is the first to report the application of Raman spectroscopy in the analysis of pharmacodynamic substances and soluble solid in the manufacturing process of Guanxinning tablets. The proposed method is also expected to be useful for the implementation of process analytical techniques in the manufacturing of other botanical drugs.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules27206969/s1>. Table S1. Calibration curves, correlation coefficients, linearity ranges, LOD and LOQ data of the four bioactive compounds. Table S2. Precision, repeatability, stability, and recovery of the four bioactive compounds. Table S3. The contents range of bioactive ingredients and soluble solid in training sets and test sets. Table S4. The performance parameters of the PLSR models established with different characteristic bands selecting methods.

Author Contributions: Conceptualization, Y.T. and L.L.; methodology, J.B. and Q.L.; software, J.Z.; writing—original draft preparation, J.B. and Y.T.; writing—review and editing, Y.T., L.L. and J.Z.; supervision, Y.T. and J.Z.; and funding acquisition, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: Financial support is gratefully acknowledged from the Natural Science Foundation of Zhejiang Province (No. Y21H280036) and Horizontal Scientific Research Project (No. KYY-HX-20220295).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and algorithm can be found at <https://github.com/zjuttaoyi/-/tree/CNN>.

Acknowledgments: The authors would like to thank the anonymous reviewers for their invaluable suggestions that helped improve the manuscript.

Conflicts of Interest: The author declares no conflict of interest.

Sample Availability: Samples of the compounds are not available from authors.

Abbreviations

CNN: convolutional neural network; CARS: competitive adaptive reweighted sampling; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross-validation; RMSEP: root mean square error of prediction; R_c^2 : coefficient of determination of calibration; R_{cv}^2 : coefficient of determination of cross-validation; R_p^2 : coefficient of determination of validation; PAT: process analysis technology; DAD: diode array detector; UVE: Uninformative Variable Elimination; SPA: Successive Projections Algorithm; siPLS: Synergy Interval Partial Least Square; LV: latent variable; PRESS: predicted residual sum squared; K-S: Kennard–Stone; S-G: Savitzky–Golay; and ReLU: Rectified linear units.

References

1. Hou, Y.; Cai, X.; Miao, P.; Li, S.; Shu, C.; Li, P.; Li, W.; Li, Z. A feasibility research on the application of machine vision technology in appearance quality inspection of Xuesaitong dropping pills. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *258*, 119787. [[CrossRef](#)] [[PubMed](#)]
2. Zhang, Z.; Zhang, Y.; Wang, L.; Cui, T.; Wang, Y.; Chen, J.; Li, W. On-line screening of natural antioxidants and the antioxidant activity prediction for the extracts from flowers of *Chrysanthemum morifolium* ramat. *J. Ethnopharmacol.* **2022**, *294*, 115336. [[CrossRef](#)] [[PubMed](#)]
3. Food and Drug Administration. *Guidance for Industry: PAT-A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance*; Food and Drug Administration: Washington, DC, USA, 2004.
4. Guo, F.L. Pharmacological effects of *Salvia miltiorrhiza* and its clinical application. *Chin. J. Hosp. Pharm.* **1999**, *19*, 363–365.
5. Wang, C.S.; Yang, J.R.; Gui, C.Q.; Ding, B.P.; Song, J.G. Effects of *Salvia miltiorrhiza* injection on acute myocardial ischemia and hemorheology models of rat. *Chin. J. Clin. Pharmacol. Ther.* **2002**, *1*, 30–32.
6. Chen, M.L.; Shou, Q.Y.; Pan, Y.M.; Zhang, J.B.; Sang, R.; Guan, M.W.; Wu, B.Y. Inhibitive and protective effect of guanxinling tablets on platelet aggregation and vascular endothelium in qi stagnation and blood stasis rats. *Chin. J. Clin. Pharmacol. Ther.* **2005**, *5*, 586–589.
7. Jurašeková, Z.; Fabriciová, G.; Silveira, L.F.; Lee, Y.N.; Gutak, J.M.; Ataabadi, M.M.; Kundrať, M. Raman Spectra and Ancient Life: Vibrational ID Profiles of Fossilized (Bone) Tissues. *Int. J. Mol. Sci.* **2022**, *23*, 10689. [[CrossRef](#)] [[PubMed](#)]
8. Alali, H.; Ai, Y.; Pan, Y.L.; Videen, G.; Wang, C. A Collection of Molecular Fingerprints of Single Aerosol Particles in Air for Potential Identification and Detection Using Optical Trapping-Raman Spectroscopy. *Molecules* **2022**, *27*, 5966. [[CrossRef](#)] [[PubMed](#)]
9. Sun, J.; Xu, X.; Feng, S.; Zhang, H.; Xu, L.; Jiang, H.; Sun, B.; Meng, Y.; Chen, W. Rapid identification of salmonella serovars by using Raman spectroscopy and machine learning algorithm. *Talanta* **2022**, *253*, 123807. [[CrossRef](#)] [[PubMed](#)]

10. Palermo, A.; Sodo, A.; Naciu, A.M.; Gioacchino, M.D.; Paolucci, A.; Masi, A.D.; Maggi, D.; Crucitti, P.; Longo, F.; Perrella, E.; et al. Clinical use of Raman spectroscopy improves diagnostic accuracy for indeterminate thyroid nodules. *J. Clin. Endocrinol. Metab.* **2022**, *107*, dgac537. [[CrossRef](#)] [[PubMed](#)]
11. Zhao, Y.; Ji, N.; Yin, L.; Wang, J. A non-invasive method for the determination of liquid injectables by Raman spectroscopy. *AAPS PharmSciTech* **2015**, *16*, 914–921. [[CrossRef](#)] [[PubMed](#)]
12. Zhao, Y.; Ma, X.; Shen, Y.T.; Nan, N.; Cao, L.M.; Yi, L.H. Non-Invasive determination of pethidine hydrochloride injection by Raman spectroscopy. *Chin. Pharm. J.* **2016**, *51*, 925–929.
13. Zhang, R.; Wu, X.; Chen, Y.; Xiang, Y.; Liu, D.; Bian, X. Grey Wolf Optimizer for Variable Selection in Quantification of Quaternary Edible Blend Oil by Ultraviolet-Visible Spectroscopy. *Molecules* **2022**, *27*, 5141. [[CrossRef](#)] [[PubMed](#)]
14. Aguilar-Hernández, I.; Afseth, N.K.; López-Luke, T.; Contreras-Torres, F.F.; Wold, J.P.; Ornelas-Soto, N. Surface enhanced Raman spectroscopy of phenolic antioxidants: A systematic evaluation of ferulic acid, p-coumaric acid, caffeic acid and sinapic acid. *Vib. Spectrosc.* **2017**, *89*, 113–122. [[CrossRef](#)]
15. Sidoryk, K.; Filip, K.; Cmoch, P.; Łaszcz, M.; Cybulski, M. Efficient synthesis and physicochemical characterization of natural danshensu, its S isomer and intermediates thereof. *J. Mol. Struct.* **2018**, *1153*, 135–148. [[CrossRef](#)]
16. Li, H.-D.; Xu, Q.-S.; Liang, Y.-Z. libPLS: An integrated library for partial least squares regression and linear discriminant analysis. *Chemometr. Intell. Lab.* **2018**, *176*, 34–43. [[CrossRef](#)]
17. Liu, J.; Osadchy, M.; Ashton, L.; Foster, M.; Solomon, C.J.; Gibson, S.J. Deep convolutional neural networks for Raman spectrum recognition: A unified solution. *Analyst* **2017**, *142*, 4067–4074. [[CrossRef](#)] [[PubMed](#)]
18. Yan, X.; Zhang, S.; Fu, H.; Qu, H. Combining convolutional neural networks and on-line Raman spectroscopy for monitoring the Cornu Caprae Hircus hydrolysis process. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *226*, 117589. [[CrossRef](#)]
19. Acquarelli, J.; van Laarhoven, T.; Gerretzen, J.; Tran, T.N.; Buydens, L.M.C.; Marchiori, E. Convolutional neural networks for vibrational spectroscopic data analysis. *Anal. Chim. Acta* **2017**, *954*, 22–31. [[CrossRef](#)]
20. Chan, C.O.; Chu, C.C.; Mok, D.K.; Chau, F.T. Analysis of berberine and total alkaloid content in cortex phellodendri by near infrared spectroscopy (NIRS) compared with high-performance liquid chromatography coupled with ultra-visible spectrometric detection. *Anal. Chim. Acta* **2007**, *592*, 121–131. [[CrossRef](#)] [[PubMed](#)]