Effect of different spectral processing methods on the quantitative model of Astragalus polysaccharides detected by NIR spectroscopy

Fan ZHAO[1] (iD), Jiawei ZHANG[1*], Jihao ZHI[1], Zhaoru YAN[1]

Abstract

Quality control of Chinese herbal medicines requires reliable and rapid detection techniques, Many scholars have used NIR spectroscopy for research in Chinese herbal medicine industry. The aim of this study was to evaluate the feasibility of combining NIR spectroscopy and machine learning to predict the adulteration content of Astragalus polysaccharides. The astragalus polysaccharide and rice flour were mixed to form astragalus polysaccharide adulterant with different concentration. the Quantitative prediction models are built using the support vector machine (SVM). Different pre-processing methods of SG, SNV, WT, SG + SNV, SG + WT were used to process the spectra, and then the continuous projection algorithm (SPA) and uninformative variable elimination (UVE) were used to select the characteristic wavelengths. The SVM prediction model based on SG + SNV + SPA is optimal with an RPD of 2.653, indicating that the model has good predictive ability. The results showed that NIR technique could be used for quantitative analysis of Astragalus polysaccharides.

Keywords: Astragalus polysaccharide; NIR; processing; SVM model.

Practical Application: Detection of adulteration of traditional Chinese medicine

1 Introduction

Astragalus is a traditional Chinese herb that has been used in China for more than 2000 years. Astragalus is mostly produced in Inner Mongolia, Shanxi, Gansu and Heilongjiang, China. It is known for its immune enhancing, anti-aging, anti-tumor, and antihypertensive effects (Wang et al., 2021a; Yu et al., 2021; Lee et al., 2020). Astragalus is known as a "good qi tonic", which has the advantages of good effect and moderate price compared with ginseng and other high-class qi tonic herbs. The main components of Astragalus are astragalus polysaccharides, saponins and flavonoids, among which astragalus polysaccharides are one of the most important natural active ingredients in Astragalus (Zheng et al., 2020). Astragalus polysaccharide (APS) is a water-soluble heteropolysaccharide extracted, concentrated and purified from the rhizome of Astragalus, which is light yellow in color and fine powder. The composition of Astragalus polysaccharide is complex. From the current analysis and research results, Astragalus polysaccharide mainly includes glucose, heteropolysaccharide, neutral polysaccharide and acid polysaccharide. Clinical studies and experiments have shown that Astragalus polysaccharide has the ability to improve the function of liver, spleen and other organs, can regulate blood sugar and blood pressure, protect blood vessels, enhance human immunity, in addition to anti-viral and anti-aging antioxidant effects (Fan et al., 2021; Wang et al., 2021b; Li et al., 2021).

Astragalus polysaccharide is in great demand in the Chinese and international markets and has a broad development prospect.

At present, there are many kinds of Astragalus polysaccharides and different prices in the Chinese market, and there are phenomena of adulterating the raw materials of Astragalus polysaccharides with substandard ones, such as adulterating rice powder and anhydrous glucose in Astragalus polysaccharides. For many years, the problem of adulteration of Astragalus polysaccharide has been interfering with the normal development of Astragalus polysaccharide market.

The existing methods for the identification of Astragalus polysaccharides are physical and chemical methods such as performance identification, solubility identification, alcoholic sedimentation method (Jiang et al., 2020; Liu et al., 2020b). These methods are deficient in that physical identification methods are artificially subjective, while chemical identification methods are destructive to Astragalus polysaccharides and the identification procedure is time-consuming and laborious. For the normal operation of the Astragalus market, it is necessary to find a fast and non-destructive method for the identification of adulterated Astragalus polysaccharides.

Near Infrared Spectrum (NIR) is a wave of electromagnetic radiation between the visible (Vis) and mid-infrared (MIR). The NIR spectral region coincides with the absorption regions of the ensemble and different multiples of the vibrations of hydrogen-containing groups (e.g., O-H, N-H, C-H) in organic molecules (Chang et al., 2020). By analyzing the NIR spectra, information on the characteristics of the hydrogen-containing groups of organic molecules in the samples can be obtained.

Near infrared spectroscopy combined with stoichiometry can be used for quantitative and qualitative analysis of products, for example, near infrared spectroscopy technology is applied to automatic identification of cheese (Silva et al., 2022). Khan et al. (2021) used near-infrared spectroscopy to analyze the quality of milk powder, and the accuracy of PL model to predict the quality of milk powder was 88%-90% (Khan et al., 2021). Silva et al. (2021) found that near-infrared spectroscopy is very useful in identifying cheese according to milk source (Silva et al., 2021).

In the field of Chinese herbal medicines, the nondestructive detection of the content of specific components of samples using NIR spectroscopy has been widely used. Certain scholars have used NIR spectroscopy to study herbs such as red ginseng, salvia miltiorrhiza, carbonized Typhae Pollen (Chen et al., 2021; Zhang et al., 2021; Ming-Liang et al., 2022).

In recent years, scholars have conducted fewer studies on Astragalus and Astragalus polysaccharides using near infrared. the study of adulteration of Astragalus polysaccharides using NIR spectroscopy has been rarely reported. In this paper, the adulterated products of Astragalus polysaccharides with different concentrations were studied using NIR spectroscopy.

2 Materials and methods

2.1 Sample preparation

2021.11.25, The Astragalus polysaccharide powder used in this paper was purchased from Evergreen Biological Engineering Co. in Shaanxi, China, as shown in Figure 1. The Certificate of Analysis (CAS) provided by this company showed that Astragalus polysaccharide was extracted from Astragalus root. The index composition of Astragalus polysaccharide was 90.0% and the result showed 91.3%. The appearance of Astragalus polysaccharide is brown fine powder, the particle size can pass 80 mesh sieve, and the bulk density is 40-60 g/100 mL.

The rice was purchased from the farmers' market in Harbin, Heilongjiang Province, China, and crushed using a multifunctional Chinese medicine high-speed grinder (model: 2500C, fineness of crushing: (50-300 mesh), and the crushed rice was passed through an 80-mesh sieve to obtain the rice powder used in this paper.

Different proportions of Astragalus polysaccharide powder and rice flour were mixed to form a mixture of rice and Astragalus polysaccharide, mixed into rice flour at the following concentrations: 0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50%. There were 60 samples for each concentration gradient and 11 concentrations, for a total of 660 samples. The astragalus polysaccharide mixtures of different concentration gradients are shown in Figure 1. The color difference between adjacent concentrations of astragalus polysaccharide mixtures was not significant and could not be distinguished by appearance.

2.2 NIR spectra acquisition

A portable NIR spectrometer from US Ocean Optics Ltd, model NIRQuest-512, with a wavelength range of 900-1700 nm, was used in reflection mode. The NIR acquisition process was controlled by Ocean View software and the diffuse reflectance standard was white. NIR spectra were collected for Astragalus polysaccharide mixture and rice powder.

2.3 Spectral preprocessing

There are various spectral pre-processing methods, which can be divided into four categories: baseline correction, scattering correction, smoothing and scale scaling. Baseline correction includes first-order derivative, second-order derivative, wavelet transform, etc.; scattering correction includes multiple scattering correction, standard normal transform, etc.; smoothing includes SG smoothing, etc.; scale scaling includes neutralization, maximum-minimum normalization, etc.

In this paper, wavelet threshold denoising (WT), standard normal variables (SNV), SG (Savitzky Golay) smoothing and SG + SNV, SG + CWT spectral preprocessing methods are used for NIR spectral preprocessing techniques.
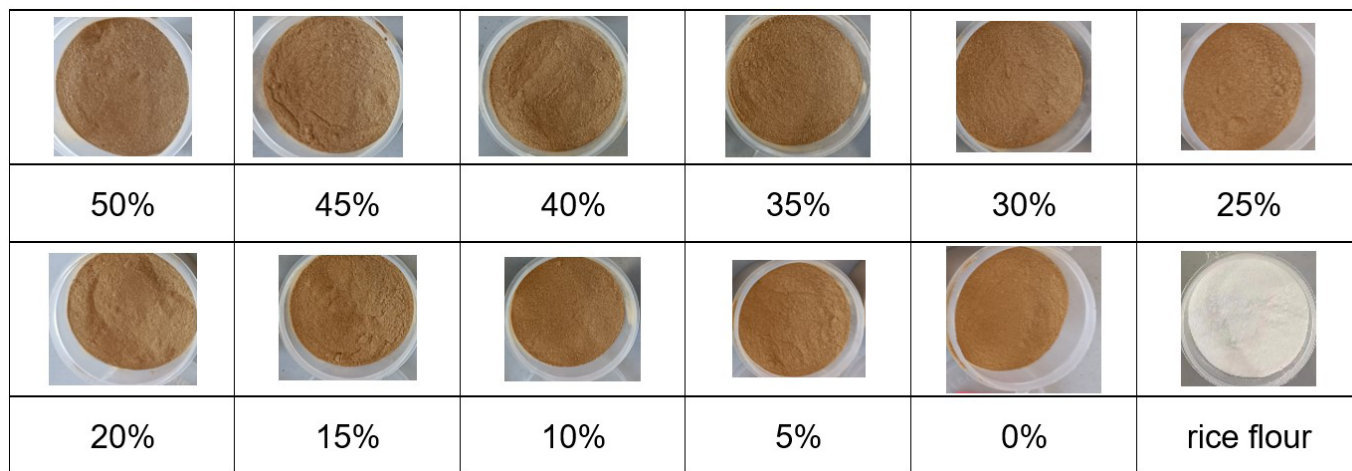

Figure 1. Astragalus polysaccharide mixture with different concentrations of rice flour.

2.4 Data analysis

Due to the large number of NIR spectral bands and the complexity of the data, the continuous projection algorithm (SPA) and the uninformative variable elimination (UVE) method are used for data dimensionality reduction. The continuous projection algorithm (SPA) is a forward variable selection algorithm that minimizes the covariance of the vector space, which can eliminate redundant information in the original spectral matrix and can be used for spectral feature wavelength screening (Zhao et al., 2021). UVE is a variable selection method based on partial least squares regression coefficients, which can eliminate information variables and improve model accuracy (Wang et al., 2022).

2.5 SVM model

SVM are a common model analysis method that excels in coping with linearly indistinguishable sample data. By using a nonlinear mapping algorithm the samples that are indistinguishable in the low-dimensional input space, are transformed into a high-dimensional feature space making them linearly distinguishable. The samples required by the SVM model are less and kernel functions are introduced.

2.6 Model judgment metrics

The correlation coefficient (R_c), root mean square of calibration set (RMSEC), correlation coefficient (R_p), root mean square of predication set (RMSEP), and residual prediction deviation (RPD) of the prediction set are used as the evaluation indexes. Prediction deviation (RPD) as the evaluation index of model performance.

A good model should have large R_c and R_p, as well as small RMSEC and RMSEP values. an RPD value below 1.5 indicates that calibration is not available. RPD values in the interval 1.5-2.0 enable to distinguish between high and low values; RPD values between 2.0-2.5 make quantitative predictions possible; for values between 2.5-3.0 and above 3.0, predictions are considered good and excellent (Tian et al., 2020; Liu et al., 2020a).

3 Results and analysis

3.1 Raw spectra of samples

The beginning and end of the original spectra were noisy, the opening and end spectra were removed, and the spectra in the range of 987-1672 nm were taken as the effective spectra, with 426 bands. The spectral curves of astragalus polysaccharide and rice powder are shown in Figure 2. These two spectral curves showed similar trends with absorption peaks at 1200 nm, 1430 and 1671 nm.

In the ranges of 987-1410 nm and 1524-1672 nm, the spectral reflectance of rice powder was greater than that of astragalus polysaccharide, while in the range of 1410-1524 nm, the reflectance of astragalus polysaccharide was transiently higher than that of rice powder.

The average spectra of different concentrations of astragalus polysaccharide mixtures are shown in Figure 3. The trends of the spectral curves were basically the same, all with a clear

trough at 1200 nm and between 1450-1650 nm. The average spectral curves of the astragalus polysaccharide mixtures with concentrations of 100%, 95% and 90% clustered together, while the spectral curves of the other concentrations clustered together, with the former having an overall higher reflectance than the latter.

3.2 Spectral pre-processing and feature spectrum extraction methods

The spectral curves after SG smoothing, SNV, WT, SG + SNV, and SG + WT pre-processing, respectively, are shown in Figure 4. The preprocessed spectra were divided into samples using the SPXY method with a correction set and prediction set ratio of 2:1, and 440 correction set samples and 220 prediction set samples were obtained.
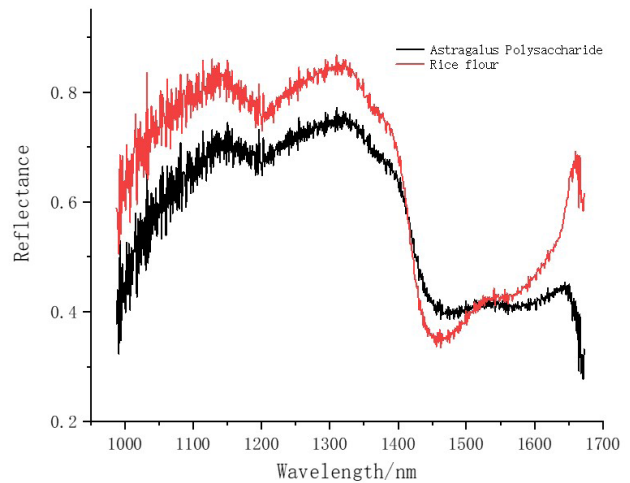

Figure 2. Spectral curves of Astragalus Polysaccharides and rice flour.
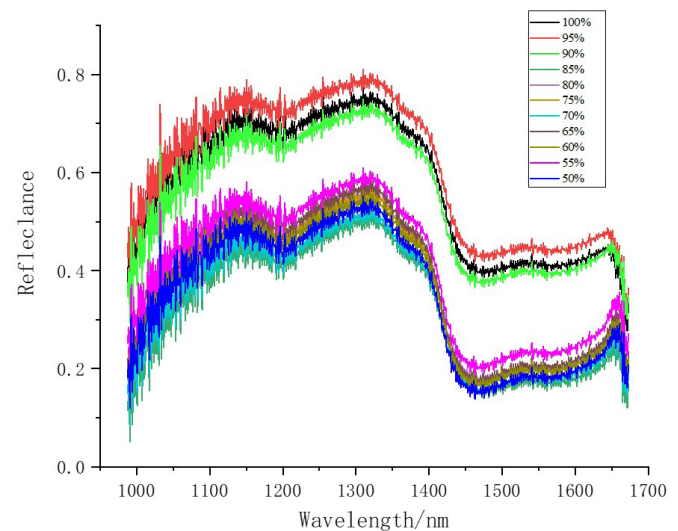

Figure 3. Average reflectance spectra of Astragalus polysaccharide mixtures with different concentrations.
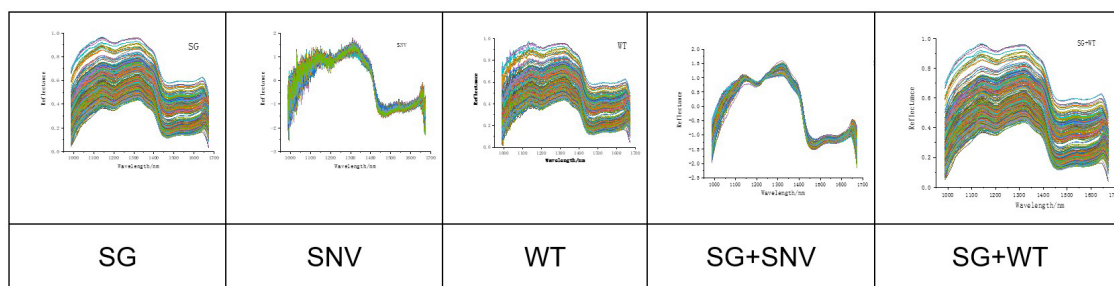
Figure 4. Spectral curve after pretreatment.

The number of feature wavelengths selected by SPA and UVE are shown in Figure 5. The overall number of feature wavelengths extracted by SPA algorithm is small, which can greatly reduce the model input; the number of feature wavelengths extracted by UVE algorithm has a wide range, with the minimum value of 7 and the maximum value is 274.

3.3 Predictive models

The SVM model uses radial basis function (RBF) as the kernel function. The K-fold Cross Validation (K-CV) method is used to determine the optimal penalty coefficient c and the kernel function parameter g. The spectra after different pre-processing methods and different feature band extraction methods are used as the input of the model to build the SVM prediction model, respectively.

SVM prediction model after SG pre-processing

The SVM prediction model after SG preprocessing is shown in Table 1, and the spectra after SG, SG + SPA, and SG + UVE were used as model inputs to build the SVM model. the SVM model with SG + SPA had the highest Rp and RPD of 0.931 and 2.559, respectively, and the lowest RMSEP of 0.054. this set of data indicates that the SVM model with SG + SNV prediction performance is good.

SVM prediction model after SNV pre-processing

The spectra after SNV, SNV + SPA, and SNV + UVE were used as model inputs to build the SVM models, respectively, and the results are shown in Table 1. The model Rp of all three is higher than 0.9, and the RPD is greater than 2.0, which indicates that the prediction is possible. Among them, the prediction ability of the SVM model of SNV is slightly better than that of SNV+SPA, but the model input of the latter is 3 wavelength numbers, which greatly reduces the calculation of the model. Considered together, the models of both are considered good.

SVM prediction model after WT pre-processing

The spectra after WT, WT + SPA, and WT + UVE were used as model inputs to build the SVM models, respectively, and the results are shown in Table 1. the RPD of the SVM models of WT and WT + SPA were less than 1.5, indicating that the prediction effect was poor; while the RPD of the SVM model of WT + UVE was 1.531 and the Rp was 0.917, indicating that this model could distinguish between high and low values.
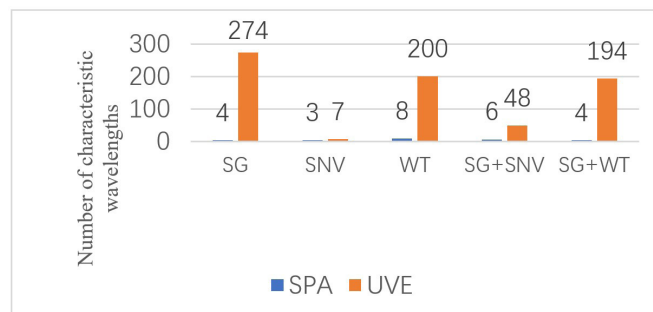

Figure 5. Number of feature wavelengths extracted by different pre-processing and feature extraction methods.

SVM prediction model after SG + SNV pre-processing

The spectra after SG + SNV, SG + SNV + SPA, and SG + SNV + UVE were used as model inputs to build the SVM models, respectively, and the results are shown in Table 1. The Rp of these three models were all greater than 0.91 and the RPD were all greater than 2.5, which indicated that the prediction ability of these three models was good. Among them, the SVM model of SG + SNV + SPA had the highest prediction RPD, which reached 2.653.

SVM prediction model after SG + WT pre-processing

The spectra after SG + WT, SG + WT + SPA, and SG + WT + UVE were used as model inputs to build the SVM models, respectively, and the results are shown in Table 1. Compared with the SVM models of SG + WT + SPA, the SVM models of SG + WT and SG + WT + UVE have lower Rp and RPD, and the model prediction ability is not good. While the SVM model of SG + WT + SPA has the highest Rp and RPD, which are 0.912 and 2.260, respectively, and can achieve prediction.

3.4 Model comparison

Comparing the prediction ability of SVM models with different preprocessing methods of the spectrum, it is found that the SVM model of WT has the worst overall prediction ability; the SVM model of SNV has better prediction ability; and the SVM model of SG+SNV has the best overall prediction ability, among which, the SVM model of SG + SNV + SPA has the best prediction ability with the highest RPD of 2.653, while the Rp is also higher, 0.939. As the optimal model, the SVM model of SG + SNV + SPA is shown in red box in Table 1.

Table 1. Spectral SVM model results based on different preprocessing processing methods.

Spectral processing method	R_c	RMSEC/Brix	R_p	RMSEP/Brix	RPD
SG	0.970	0.040	0.819	0.157	0.872
SG + SPA	0.960	0.047	0.931	0.054	2.559
SG + UVE	0.965	0.044	0.880	0.122	1.123
SNV	0.979	0.032	0.949	0.072	2.355
SNV + SPA	0.885	0.071	0.942	0.072	2.345
SNV + UVE	0.873	0.075	0.910	0.083	2.040
WT	0.975	0.037	0.858	0.127	1.092
WT + SPA	0.959	0.047	0.831	0.152	0.912
WT + UVE	0.964	0.044	0.917	0.090	1.531
SG + SNV	0.957	0.047	0.916	0.060	2.486
SG + SNV + SPA	0.936	0.057	0.939	0.056	2.653
SG + SNV + UVE	0.947	0.052	0.944	0.062	2.393
SG + WT	0.967	0.042	0.862	0.133	1.029
SG + WT + SPA	0.958	0.047	0.912	0.061	2.260
SG + WT + UVE	0.964	0.044	0.882	0.121	1.129

4 Conclusion

NIR spectra can be used to identify and quantify adulterated mixtures of Astragalus polysaccharides at different concentrations. In this paper, five spectral pre-processing methods, SG, SNV, WT, SG + SNV, SG + WT, and two feature spectral extraction methods, SPA and UVE, were used to develop a quantitative prediction model of Astragalus polysaccharide SVM. NIR spectroscopy has been widely used in the food field, but it does not provide good results in identifying and quantifying adulteration of Astragalus polysaccharides, and this study can fill this gap.

Conflict of interest

The author declares that they have no financial interests and personal relationship competition about this paper.

Acknowledgements

References

Chang, S., Yin, C., Liang, S., Lu, M., Wang, P., & Li, Z. (2020). Confirmation of brand identification in infant formulas by using near-infrared spectroscopy fingerprints. *Analytical Methods*, 12(19), 2469. http://dx.doi.org/10.1039/D0AY00375A. PMid:32930236.

Chen, B., Zheng, E.-R., & Guo, T. (2021). Application of various algorithms for spectral variable selection in NIRS modeling of red ginseng extraction. *Guangpuxue Yu Guangpu Fenxi*, 41(8), 7. http://dx.doi.org/10.3964/j.issn.1000-0593(2021)08-2443-07.

Fan, X. H., Li, K., Yang, Y. D., Qin, X. M., Li, Z. Y., Li, X.Q. (2021). Preliminary study on structure and activity of anti-inflammatory components in Astragalus Polysaccharides. *Journal of the Shanxi Medical University*, 52(10), 11. http://dx.doi.org/10.13753/j.issn.1007-6611.2021.10.016.

Jiang, X. H., Lin, N., Liu, D. M., Li, D. P., Lu, F. L. (2020). Comparative study on the quality of Astragalus Polysaccharide from different sources. *Guangxi Science*, 27(4), 7. http://dx.doi.org/10.13656/j.cnki.gxkx.20200924.002.

Khan, A., Munir, M. T., Yu, W., & Young, B. R. (2021). Near-infrared spectroscopy and data analysis for predicting milk powder quality attributes. *International Journal of Dairy Technology*, 74(1), 235-245. http://dx.doi.org/10.1111/1471-0307.12734.

Lee, Y. C., Huang, H. T., Chang, C. D., Chen, C. T., Lin, T. Y., Yang, T. W., Chang, F. L., Lu, M. K., Chiou, C. T., Chen, W. C., Pan, S. L., & Tsai, K. C. (2020). Isolation of anti-VEGF monoclonal antibodies with neutralizing effects from an Astragalus-induced immune antibody library. *International Immunopharmacology*, 88, 107007. http://dx.doi.org/10.1016/j.intimp.2020.107007. PMid:33182041.

Li, J., Jiang, L. L., Li, M. X., Deng, F., Xu, S. Y., Xue, X., Hu, R. P., Xue, H. T. (2021). Inhibitory Effects of Astragalus Polysaccharide on Activity of Xanthine Oxidase. *Journal of Food Science and Biotechnology*, 40(10), 5. http://dx.doi.org/10.3969/j.issn.1673-1689.2021.10.003.

Liu, J., Xie, J., Han, J., Wang, H., Sun, J., Li, R., & Li, S. (2020a). Visible and near-infrared spectroscopy with chemometrics are able to predict soil physical and chemical properties. *Journal of Soils and Sediments*, 20(7), 2749-2760. http://dx.doi.org/10.1007/s11368-020-02623-1.

Liu, W. B., Yu, X., Xu, J. J., Zhan, X. B., Zhang, H. T., Zhu, L. (2020b). Isolation,structure characterization and prebiotic activity of polysaccharides from Astragalus membranaceus. *Shipin Yu Fajiao Gongye*, 46(7), 50-56. http://dx.doi.org/10.13995/j.cnki.11-1802/ts.022967.

Ming-Liang, G., Yi, Z., Fang-Fang, C., Hang-Hang, W., Ling-Run, L., Xin, J., Ya-Nan, Z., Tian-Shu, W., Pei-Dong, C., Wei-Feng, Y., Bei-Hua, B., & Li, Z. (2022). A gradient-based discriminant analysis method for process quality control of carbonized TCM via Fourier transform near infrared spectroscopy: a case study on carbonized Typhae Pollen. *Spectrochimica*

Acta. Part A: Molecular and Biomolecular Spectroscopy, 265, 120363. http://dx.doi.org/10.1016/j.saa.2021.120363. PMid:34562862.

Silva, L. K. R., Jesus, J. C., Onelli, R. R. V., Conceição, D. G., Santos, L. S., & Ferrão, S. P. B. (2021). Discriminating Coalho cheese by origin through near and middle infrared spectroscopy and analytical measures. Discrimination of Coalho cheese origin. *International Journal of Dairy Technology*, 74(2), 393-403. http://dx.doi.org/10.1111/1471-0307.12767.

Silva, L. K. R., Santos, L. S., & Ferrão, S. P. B. (2022). Application of infrared spectroscopic techniques to cheese authentication: a review. *International Journal of Dairy Technology*, 75(3), 490-512. http://dx.doi.org/10.1111/1471-0307.12859.

Tian, A., Fu, C., Xiong, H., Zhao, J.S. (2020). Quantitative analysis of 2 SO4 – in saline soil under areas disturbed and undisturbed by human using BP neural network. *Chinese Journal of Eco-Agriculture*, 28(2), 296-304. http://dx.doi.org/10.13930/j.cnki.cjea.190700.

Wang, X. J., Ding, L. M., Wei, H. Y., Jiang, C. X., Yan, Q., Hu, C. S., Jia, G. X., Zhou, Y. Q., Henkin, Z., & Degen, A. A. (2021a). Astragalus membranaceus root supplementation improves average daily gain, rumen fermentation, serum immunity and antioxidant indices of Tibetan sheep. *Animal*, 15(1), 100061. http://dx.doi.org/10.1016/j.animal.2020.100061. PMid:33516026.

Wang, Y., Yang, Z. G.,Wang, Y. H., Li, J., Wang, D., Li, L. W. (2021b). Research progress on immune regulation of Astragalus Polysaccharide and ginseng polysaccharides. *Shipin Anquan Zhiliang Jiance Xuebao*, 12(3), 7. http://dx.doi.org/10.19812/j.cnki.jfsq11-5956/ts.2021.03.010.

Wang, Z., Chen, J., Zhang, J., Tan, X., Raza, M. A., Ma, J., Zhu, Y., Yang, F., & Yang, W. (2022). Assessing canopy nitrogen and carbon content in maize by canopy spectral reflectance and uninformative variable elimination. *The Crop Journal*, 10(5), 1224-1238. http://dx.doi.org/10.1016/j.cj.2021.12.005.

Yu, J., Dong, X., Jiao, J., Ji, H., & Liu, A. (2021). Antitumor and immunoregulatory activities of a novel polysaccharide from Astragalus membranaceus on S180 tumor-bearing mice. *International Journal of Biological Macromolecules*, 189, 930-938. http://dx.doi.org/10.1016/j.ijbiomac.2021.08.099. PMid:34419546.

Zhang, S., Wang, D., Zhu, J., Li,C.Y., Wang, Z. M., Kuang, Y. H. (2021). Process quality control of Tanshinones in Salvia miltiorrhiza within the extraction and concentration of Fufangdanshen Tabletsby near infrared spectroscopy. *Ginseng Research*, 33(03), 15-20. http://dx.doi.org/10.19403/j.cnki.1671-1521.2021.03.004.

Zhang, S., Wang, D., Zhu, J., Li, C. Y., Wang, Z. M., Kuang, Y. H. (2021). Hyperspectral used to recognize black goji berry and Nitraria tanggu. *Guangpuxue Yu Guangpu Fenxi*, 41(7), 2240-2244. http://dx.doi.org/10.3964/j.issn.1000-0593(2021)07-2240-05.

Zheng, Y., Ren, W., Zhang, L., Zhang, Y., Liu, D., & Liu, Y. (2020). A review of the pharmacological action of Astragalus Polysaccharide. *Frontiers in Pharmacology*, 11, 349. http://dx.doi.org/10.3389/fphar.2020.00349. PMid:32265719.