ADULTERATION IDENTIFICATION OF ASTRAGALUS POLYSACCHARIDES BY NIR SPECTROSCOPY COMBINED WITH SIMCA AND PLS-DA
/
近红外光谱结合 SIMCA 和 PLS-DA 鉴别黄芪多糖掺假

Zhao FAN, Zhang JIAWEI *, Zhi JIHAO [1]
College of mechanical and Electronic Engineering, Northeast Forestry University, Haerbin/China
Tel: +86-0450-82190397; E-mail: zjw@nefu.edu.cn

DOI: https://doi.org/10.35633/inmateh-68-82

***Keywords**: Astragalus polysaccharide, Classification, SIMCA, PLS-DA*

ABSTRACT

As a famous Chinese traditional medicine, the Astragalus polysaccharide (APS) market is continually expanding, while the quality of APS cannot be guaranteed. Near-infrared (NIR) spectroscopy has been widely used in the detection of Chinese herbal medicines and traditional Chinese medicine. In this study, NIR spectroscopy was used to identify the adulterants of APS. Prepare adulterated mixtures of APS with 75%, and 50% content, respectively. PLS-DA and SIMCA models were developed for 2-classification of APS, APS mixture (75%+50%), and 3-classification of APS, 75% APS mixture and 50% APS mixture, respectively. In the 2-classification, the correct classification rate of both the calibration set and the test set of the PLS-DA and SIMCA models is 100%. In the 3-classification, the correct classification rates of calibration set and test set for PLS-DA were 97.5% and 96.67%, respectively; the correct classification rates of calibration set and test for SIMCA were 98.33% and 100%, respectively. The study showed that it is feasible to identify adulterated Astragalus polysaccharides using near-infrared spectroscopy.

摘要

黄芪多糖（APS）作为著名的中药，其交易市场越来越大，但其质量却无法保证。近红外光谱已广泛应用于中草药的检测。在这项研究中，近红外光谱被用于鉴别 APS 的掺杂物。制备含量分别为 75%和 50%的掺杂 APS 混合物。PLS-DA 和 SIMCA 模型分别用于 APS、APS 混合物（75%+50%）的 2 级分类和 APS、75%APS 混合物和 50%APS 混合物的 3 级分类。在 2 级分类中，PLS-DA 和 SIMCA 模型的校准集和测试集的正确分类率均为 100%。在 3 级分类中，PLS-DA 的校准集和测试集的正确分类率分别为 97.5%和 96.67%；SIMCA 校准集和测试的正确分类率分别为 98.33%和 100%。研究表明，利用近红外光谱法鉴别掺假黄芪多糖是可行的。

INTRODUCTION

Astragalus polysaccharide (APS) is one of the important active ingredients of *Astragalus (Jin et al., 2014; Yan et al., 2016)*. APS is a brownish-yellow powder with hygroscopic properties, and it is a complex mixture of structure, consisting of hexose acetate, glucose, fructose and other groups. *Astragalus polysaccharide* has antiviral, antitumor, antioxidant and immune enhancing effects *(Fan et al., 2021; Li et al., 2020)*. Yu-ping Zhu et al. found that APS may be used to treat and prevent endothelial cell injury-related diseases *(Zhu et al., 2013)*; Yan-Feng Huang investigated the repairing effect of APS on fatigue, concluding that APS can improve the molecular mechanism of mitochondria *(Yan et al., 2016)*; Kojo Agyemang et al. explored the effect of APS on diabetes *(Kojo et al., 2013)*.

APS has good health effects on the organism, and the trading market of APS is gradually developing and growing. Due to the high requirements of APS for the separation and extraction process, the cost of APS is also high. The price range of APS in the market fluctuates widely, with low price of 10 Yuan/Kg and high price of 300/Kg, Low-priced APS may be adulterated with other substances, such as maltodextrin and rice flour. The common identification methods of APS are physical methods such as taste, color, and dissolution *(Jiang et al., 2020)*, which have large subjective factors and cannot be used as formal judgment criteria in the market; there are also chemical methods such as alcohol precipitation, which are lossy identification and time-consuming. Therefore, it is very important to find a fast, convenient and non-destructive method to identify APS.

[1] *Fan Zhao, AS Lec, Ph.D. Eng.; Jiawei Zhang, Prof. Ph.D.Eng.*

Near-infrared spectroscopy combined with chemometrics can enable nondestructive analysis of many herbal medicines. Linhong Fan et al. used portable NIR spectrometer to evaluate the quality of *Fritillariae cirrhosae (Fan et al., 2022)*, the identification accuracy of sources of different varieties of Fritillariae cirrhosae and the common counterfeits of *Fritillariae cirrhosae* reached 83.33% and 90.91%, respectively. Le Wang et al. used NIR spectroscopy for qualitative and quantitative analysis of *Curcumae Radix* from four botanical origins *(Le et al., 2021)*, models of support vector machine (SVM) and k-nearest neighbors (KNN) achieved the complete discrimination of the four species of Yujin with 100% accuracy. Wenlong Li et.al used near infrared spectroscopy to study the feasibility of detection in the production of traditional Chinese medicine *(Li et al., 2016).* The results showed that NIR spectroscopy was helpful to diagnose the faulty batches in the production of traditional Chinese medicine. However, there are very few studies on the identification of adulterated APS using NIR spectroscopy. This paper attempted to classify APS adulterants based on NIR spectroscopy using SIMCA (Soft Independent Modeling Class Analog) and PLS-DA (Partial least squares Discriminant Analysis).

MATERIALS AND METHODS

Samples

APS was purchased from Evergreen Bioengineering Co., Ltd. of Shanxi Province, China, which provided an analysis report showing that APS was extracted from the roots of Astragalus, passed through 80 mesh sieves, and the index composition of APS was 90.0%. Rice was purchased from a Supermarket in Harbin, Heilongjiang Province, China, it was ground and passed through 80 mesh sieves to obtain the required rice powder.

As the content of APS in adulterants is less than 50%, they can be distinguished by color without the help of NIR spectroscopy; while in the range of 50%-100%, it is impossible to distinguish by eye, which requires the use of NIR spectroscopy to distinguish. Therefore, the content of APS in the adulterants in this paper is only considered between 50% and 100%. APS was mixed with rice powder to form 3 concentrations of APS adulterants, with APS concentrations of 100%, 75% and 50%. The number of samples for each concentration was 60, with a total of 180 samples. The 100% APS was used as category A; the APS mixed with rice flour was used as category B, where the mixture with 75% APS content was used as category B1 and the mixture with 50% APS content was used as category B2. These three categories of APS mixtures are shown in Figure 1, and they are not easily discernible by external appearance.

Category A（100%） Category B1（75%） Category B2（50%）

Fig. 1 - Mixture of *Astragalus Polysaccharides* and rice flour in different concentrations

Spectral acquisition

A portable NIR spectrometer (model: NIRQuest-512) in the wavelength range of 900-1700 nm was used to collect the NIR reflectance spectra of APS adulterants.

Spectral data processing

The K-S (Kennard-Stone) algorithm is commonly used for the partitioning of data sets. The K-S algorithm is based on the principle that all samples are considered as training set candidates, from which the samples are selected in turn to enter the training set. First, the two samples with the farthest Euclidean distance are selected to enter the training set, and then, by calculating the Euclidean distance of each remaining sample to each known sample in the training set, the two samples with the farthest and closest Euclidean distances are found, and these two samples are selected to enter the training set, and the above steps are repeated until the number of samples reaches the requirement *(Morais et al., 2019; Zhang et al., 2017)*. In this study, the K-S algorithm was used to divide the samples of the three categories into calibration set and test set, respectively, in a ratio of 2:1. The number of calibrations set samples for each category was 40 and the number of test set samples was 20.

Because of the large noise at the beginning and end of the acquired spectra, the opening and end bands were removed and the band in the range of 997-1657 nm was used.

Models

SIMCA is a classification method based on Principal Component Analysis (PCA) proposed by Swedish chemists in 1976. It uses a priori classification knowledge to build a PCA model for each category, calculates the distance between the unknown sample and the PCA model, and performs discriminant analysis based on the distance *(Svante et al., 1977)*. The PCA plays a decisive role in the results during the calculation of the SIMCA method.

PLS-DA is often used for classification and discrimination. It is similar to PCA, but PCA is unsupervised analysis, while PLS-DA is supervised analysis *(Abdo et al., 2020)*. The unsupervised analysis method can distinguish between-group differences well when the differences between sample groups are large but within-group differences are small; when the differences between groups are small, the unsupervised analysis method is difficult to distinguish between-group differences. These problems can be well solved by supervised analysis like PLS-DA, which performs modeling under the condition that the number of sample points is allowed to be less than the number of variables.

SIMCA 14.1 software was applied to establish SIMCA and PLS-DA model for 2-classification analysis of A, (B1+B2); 3-classification analysis of A, B1, B2.

RESULTS AND ANALYSIS

Spectral

The average spectral curves of APS mixtures are shown in Fig. 2. It can be seen that the average spectral reflectance of class A is higher than that of classes B1 and B2 overall, while the average spectral curves of the latter two are very close. The overall trend of the A, B1, and B2 spectra is consistent, with a very small absorption near 1200 nm and a larger absorption in the 1450-1650 nm range. It is difficult to visualize the differences between species based on the full spectra of the 3 classes. Therefore, it is necessary to combine chemometric methods to develop classification models for mixtures of different concentrations of APS.

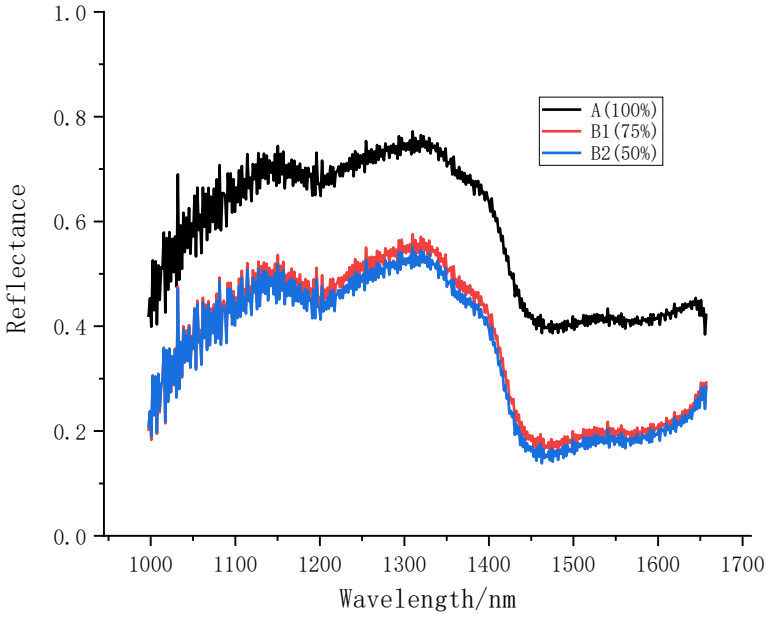

Fig. 2 - Average spectral curves of APS mixtures

2-classification models of A and B

(1) PCA analysis

Twenty samples were extracted from the calibration sets of B1 and B2, respectively, and put together with 40 calibration set samples of class A, as the calibration set of the 2-classification model, and the number of calibrations set samples was 80.

Similarly, the samples of the test set were extracted according to the above method, and the number of test set samples was 40.

All 120 samples of the 2-classifications were subjected to PCA analysis, and R2X (cum) and Q2 (cum) were both 0.995 when the first 2 principal components were taken. R2X (cum) represents the cumulative value of explained variance at the current number of principal components, while Q2 (cum) indicates the total model predictive power at the current number of principal components (*Zhang et al., 2018*). The first 2 principal component score plots are shown in Figure 4(a). The aggregation of the samples in category A is poor, while the aggregation of the samples in category B is better. However, there are individual samples confounded.

(2) PLS-DA Model

The PLS-DA model was built using SIMCA software with 7-fold cross validation (7-fold cross validation), and the values of R2Y (cum) and Q2 (cum), which represent the explanatory degree of the categorical variable Y and the predictive power of the model, respectively *(Du et al., 2021)*, were 0.991 and 0.976. The principal components of the PLS-DA model was 5.

The model was internally validated by the permutation test (permutation test), which randomly changed the order of the categorical variable Y 200 times to obtain different random Q2 values, and the permutation test plot is shown in Figure 3(b). It can be seen that all the points on the left side of R2 and Q2 are lower than those on the right side, and the regression slopes of R2 and Q2 are positive.

VIP (Variable important in projection) is the variable weight value of PLS-DA model variables, which can be used to measure the intensity and explanatory energy of the impact of the differences in the accumulation of each metabolite on the classification discriminations of each group of samples. It is generally considered that variables with VIP>1 may be differential metabolites of the two groups *(Liu et al., 2021)*, and the distribution of VIP predicted values of PLS-DA model is shown in Figure 3(a).

The results of the classification analysis of the correction and prediction sets using the established model are shown in Table 1. The PLS-DA model classifies both the correction and prediction sets with 100% correctness.

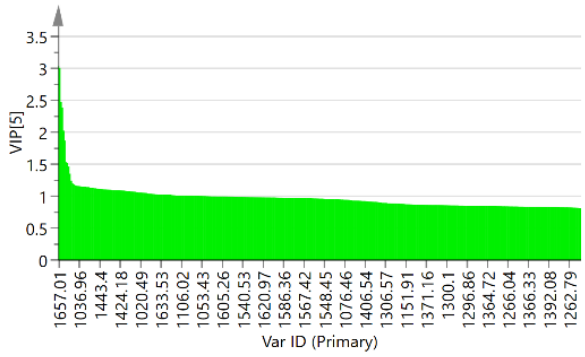

Fig. 3(a) - VIP forecast distribution

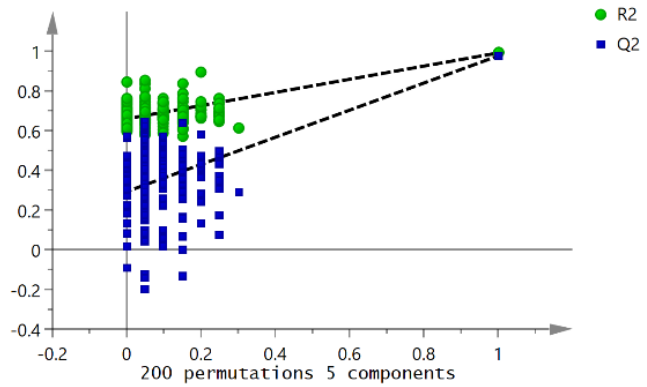

Fig. 3(b) - PLS-DA permutation test

(3) SIMCA Model

SIMCA classification model was established using SIMCA software. Firstly, the PCA models for each class of samples were established separately, and the selected principal component scores were 2 and 3 respectively; then the distances between the predicted samples and the PCA models were calculated, and the predicted samples were classified according to the distances. The results are shown in Table 1. The model classifies 100% of the calibration set and test set samples correctly, and the classification is completely correct.

Table 1

2-Classification results of APS mixtures by SIMCA and PLS-DA

Model	Samples	Class	Members	Correct	A	B
PLS-DA	Correction set	A	40	100%	40	0
		B	40	100%	0	40
		total	80	100%	100%	100%

Table 1

(continuation)

Model	Samples	Class	Members	Correct	A	B
	Test set	A	20	100%	20	0
		B	20	100%	0	20
		Total	40	1000%	20	20
SIMCA	Correction set	A	40	100%	40	0
		B	40	100%	0	40
		total	80	100%	40	40
	Test set	A	20	100%	20	0
		B	20	100%	0	20
		total	20	100%	20	20

3-Classification model of A, B1, B2

All 180 samples of the 3 categories were subjected to PCA analysis, and the R2X (cum) and Q2 (cum) of the first two principal components were both 0.995, and the plots of the scores of the first 2 principal components are shown in Figure 4(b). It can be seen that category A samples are loose and category B1 and B2 samples are more aggregated and have overlapping parts.

The number of samples in the calibration set is 120, which is used to build the SIMCA classification model, and the number of samples in the test set is 60, which is used to verify the reliability of the model.

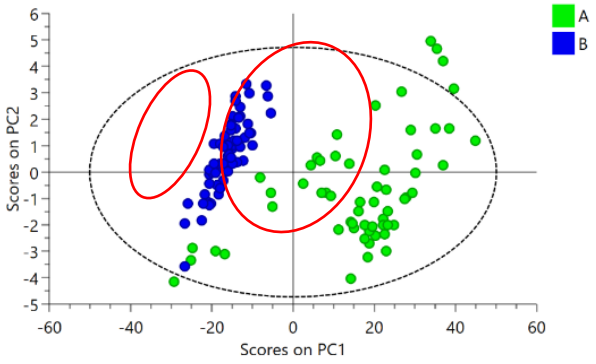

Fig. 4(a) - Principal component score chart

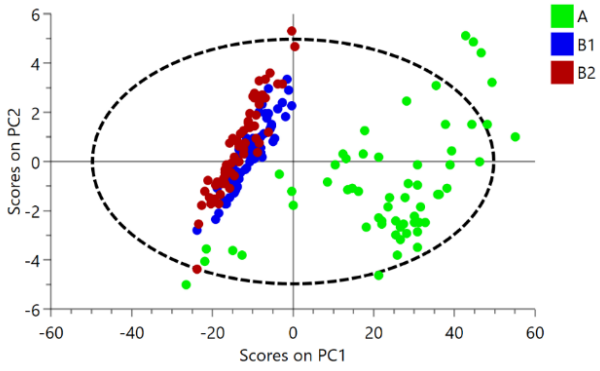

Fig. 4(b) Principal component score chart

(1) PLS-DA model

The PLS-DA classification model was established by using SIMCA software with a master score of 5, and R2Y (cum) and Q2 (cum) were both 0.86 and 0.767, respectively. The replacement test was randomized 200 times, and all the R2 and Q2 on the left side of the obtained replacement test plot were lower than the right side, and the slope was positive, see Figure 5(b). The VIP prediction distribution is shown in Figure 5(a).

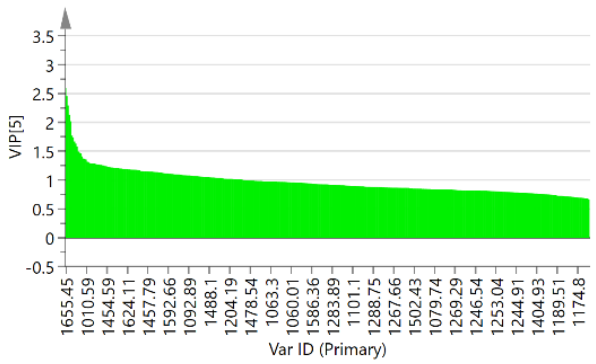

Fig. 5 (a) - VIP forecast distribution

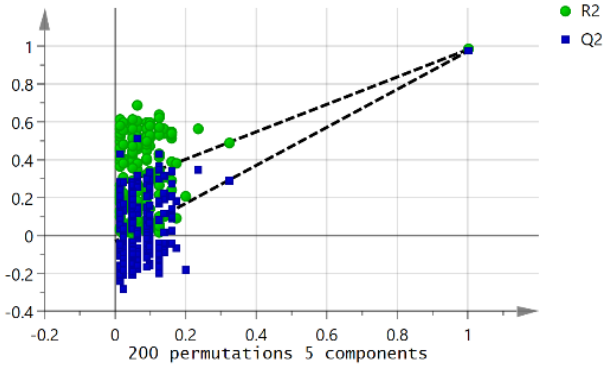

Fig. 5(b) - PLS-DA permutation test

The results of the classification analysis of the calibration set and the prediction set using the established model are shown in Table 2.

When the model classifies the calibration set, 40 samples in A are correctly classified with 100% correct rate; 39 samples in B1 are correctly classified and 1 sample in B1 is classified in B2 with 97.5% correct rate; 38 samples in B2 are correctly classified and 2 samples in B2 are incorrectly classified in B1 with 95% correct rate.

When the model classifies the test set, the 100% of the samples in A were classified correctly. 18 samples in B1 were classified correctly and 2 samples were incorrectly classified in B2, with a 90% correct rate. The samples in B2 were classified correctly with 100%.

(2) SIMCA Model

SIMCA software was used to build A, B1, and B2 SIMCA 3-classification models, and the main score of all three classification models was 2. The classification results of the models on the calibration and test sets are shown in Table 2.

In the calibration set classification results, all 40 samples in A were correctly classified with a 100% correct rate; all 40 samples in B1 were correctly classified with a 100% correct rate; 38 samples in B2 were correctly classified and 2 samples in B2 were incorrectly classified in B1 with a 95% correct rate.

In the test set classification results, all samples A, B1 and B2 were correctly classified with 100% correct rate.

Table 2

3-Classification results of APS mixtures by SIMCA and PLS-DA

Model	Samples	Class	Members	Correct	A	B1	B2
PLS-DA	Correction set	A	40	100%	40	0	0
		B1	40	97.5%	0	39	1
		B2	40	95%	0	2	38
		total	120	97.5%	40	41	39
	Test set	A	20	100%	20	0	0
		B1	20	90%	0	18	2
		B2	20	100%	0	0	20
		total	60	96.67%	20	18	22
SIMCA	Correction set	A	40	100%	40	0	0
		B1	40	100%	0	40	0
		B2	40	95%	0	2	38
		total	120	98.33%	40	42	38
	Test set	A	20	100%	20	0	0
		B1	20	100%	0	20	0
		B2	20	100%	0	0	20
		total	60	100%	20	20	20

Model Comparison

Comparing PLS-DA model with SIMCA model, the results are as follows:

(1) For A and B classification, the correctness of PLS-DA correction set and test set is 100%; The correctness of SIMCA calibration set and test set also reach 100%. PLS-DA and SIMCA performed extremely well in A and B classification, with 100% correct classification rates.

(2) For the A, B1 and B2 classifications, the correction set correctness of PLS-DA and SIMCA is 97.5% and 98.33% respectively; and the test set correctness of PLSA-DA and SIMCA is 96.67% and 100% respectively. SIMCA performed better than PLS-DA in A, B1 and B2 classification.

CONCLUSIONS

This study showed that NIR spectroscopy combined with PLS-DA and SIMCA models can identify APS mixtures. This method has the advantages of simple operation, of being non-destructive and rapid.

(1) APS (100%) and APS mixture (containing both 75% and 50%) were very easy to classify, and the average correctness of the calibration set and test set for all 2-classifition models is 100%.

(2) When the mixture with APS content of 100%, 75% and 50% is classified into 3 categories, the SIMCA model can completely classify the test set correctly, with a correct rate of 100%; while PLS-DA wrongly considered that the two samples with 50% APS content in the test set had 75% APS content, which was correct at 96.67%.

(3) In the 2-classification, SIMCA was as good as PLS-DA, with a 100% correct rate, while in the 3-classification, SIMCA was 100% correct, which was higher than PLS-DA (96.67%), the SIMCA model performs better than PLS-DA.

DECLARATIONS

Fundings This work was supported by the National Key R&D Program of China (Project No. 2020YFC2200800).

Ethics approval (include appropriate approvals or waivers): This article does not include any ethical research on humans or animals.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] Abdo H., (2020). Performance of Fluorescence and Diffuse Reflectance Hyperspectral Imaging for Characterization of Lutefisk: A Traditional Norwegian Fish Dish. *Molecules*, vol. 25, no. 1191, Switzerland, doi: 10.3390/molecules25051191.

[2] Du Y., (2021). Serum Metabolomics Study of Papillary Thyroid Carcinoma Based on HPLC-Q-TOF-MS/MS. *Frontiers in Cell and Developmental Biology*, vol. 9, no. February, pp. 1–13, Switzerland, doi: 10.3389/fcell.2021.593510.

[3] Fan L.H., (2022). Quality assessment of Fritillariae cirrhosae using portable NIR spectrometer. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, vol. 265, p. 120325, England, doi: 10.1016/j.saa.2021.120325.

[4] Fan Y.P., (2012). Effects of Astragalus polysaccharide liposome on lymphocyte proliferation in vitro and adjuvanticity in vivo. *Carbohydrate Polymers*, vol. 88, no. 1, pp. 68–74, England, doi: 10.1016/j.carbpol.2011.11.067.

[5] Jiang X.H., (2020). Comparative Study on the Quality of Astragalus Polysaccharide from Different Sources (不同来源的黄芪多糖产品质量比较研究). *Guangxi Sci.*, vol. 27, pp. 380–386, Guangxi/P.R.C., doi: 10.13656/j.cnki.gxkx.20200924.002.

[6] Jin M.L., (2014). Structural features and biological activities of the polysaccharides from Astragalus membranaceus. *International Journal of Biological Macromolecules*, vol. 64, pp. 257–266, Netherlands, doi: 10.1016/j.ijbiomac.2013.12.002.

[7] Kojo A., (2013). Recent advances in Astragalus membranaceus anti-diabetic research: Pharmacological effects of its phytochemical constituents. *Evidence-based Complementary and Alternative Medicine*, vol. 2013, England, doi: 10.1155/2013/654643.

[8] Le W. et al, (2021). Fast discrimination and quantification analysis of Curcumae Radix from four botanical origins using NIR spectroscopy coupled with chemometrics tools. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy,* vol. 254, p. 119626, England, doi: 10.1016/j.saa.2021.119626.

[9] Li W.F., (2020). Characterization and anti-tumor bioactivity of Astragalus polysaccharides by immunomodulation. *International Journal of Biological Macromolecules,* vol. 145, pp. 985–997, Netherlands, doi: 10.1016/j.ijbiomac.2019.09.189.

[10] Li W.L. et al, (2016). A feasibility research on the monitoring of traditional Chinese medicine production process using NIR-based multivariate process trajectories. *Sensors and Actuators B-Chemical*, vol. 231, pp. 313–323, Switzerland, doi: 10.1016/j.snb.2016.03.023.

[11] Liu L.C., (2021). Characterization and identification of different Chinese fermented vinegars based on their volatile components. *Journal of Food Biochemistry*, vol. 45, no. 3, pp. 1–12, United States, doi: 10.1111/jfbc.13670.

[12] Morais C. et al, (2019). Improving data splitting for classification applications in spectrochemical analyses employing a random-mutation Kennard-Stone algorithm approach. *Bioinformatics*, vol. 35, no. 24, pp. 5257–5263, England, doi: 10.1093/bioinformatics/btz421.

[13] Svante W. & Michael S., (1977). SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy. *Chemometrics: Theory and application,* ISBN: 9780841203792.
[14] Yan F.H., (2016). Effects of Astragalus Polysaccharides on Dysfunction of Mitochondrial Dynamics Induced by Oxidative Stress. *Oxidative Medicine and Cellular Longevity,* vol. 2016, United States, doi: 10.1155/2016/9573291.
[15] Zhang L.N., (2017). Kennard-Stone combined with least square support vector machine method for noncontact discriminating human blood species. *Infrared Physics and Technology*, vol. 86, pp. 116–119, Netherlands, doi: 10.1016/j.infrared.2017.08.020.
[16] Zhang X., (2018). Determination of bitterness of Andrographis herba based on electronic tongue technology and discovery of the key compounds of bitter substances. *Molecules*, vol. 23, no. 12, pp. 1–13, Switzerland, doi: 10.3390/molecules23123362.
[17] Zhu Y.P., (2013). Astragalus polysaccharides suppress ICAM-1 and VCAM-1 expression in TNF-α-treated human vascular endothelial cells by blocking NF-κB activation. *Acta Pharmacologica Sinica,* vol. 34, no. 8, pp. 1036–1042, Shanghai/P.R.C., doi: 10.1038/aps.2013.46.