Original Article
Food Science and Technology

ISSN 0101-2061 (Print)
ISSN 1678-457X (Online)

DOI: https://doi.org/10.1590/fst.98922



# Fast and nondestructive discrimination of fresh tea leaves at different altitudes based on near infrared spectroscopy and various chemometrics methods

Qinghai JIANG[1], Song MEI[1], Caixue ZHAN[1], Caihong REN[1], Zhiyu SONG[1*], Shengpeng WANG[2*]

**Abstract**

Near infrared spectroscopy (NIRS) combined with various chemometrics methods was tried to identify the fresh tea leaves at different altitudes quickly and nondestructively. Three kinds of samples were collected, then scanning NIRS, conducting spectral preprocessing to remove noise information, using backward interval partial least squares to screen characteristic spectral intervals, going on principal component analysis, respectively. Finally, least squares support vector machine method (LS-SVM) was applied to establish NIRS models, whose robustness was tested by prediction set samples. The best pretreated method was the combination of multivariate scattering correction and the first derivative. Six characteristic spectral intervals were screened, and the corresponding spectral wavenumbers were 4821.2-5091.2 $cm^{-1}$, 5368.9-5638.8 $cm^{-1}$, 6190.4-6460.4 $cm^{-1}$, 7011.9-7281.9 $cm^{-1}$, 8924.9-9191.1 $cm^{-1}$ and 9734.9-10000 $cm^{-1}$. The cumulative contribution rate of the first three principal components was 99.92%. The root mean square error of the cross validation and the determination coefficient of the calibration set model were 0.027 and 0.973, respectively. The root mean square error and the determination coefficient of the prediction set model were 0.034 and 0.968, respectively. The discrimination accuracy in prediction set was 100%. The results showed NIRS combined with LS-SVM can realize fast and nondestructive discrimination of fresh tea leaves at different altitudes.

**Keywords:** fresh tea leaves; altitude; near infrared spectroscopy; backward interval partial least squares; least squares support vector machine.

**Practical Application:** Fast discrimination of fresh tea leaves at different altitudes.

## 1 Introduction

The quality of fresh tea leaves is the basis of the quality of processed tea. Only high-quality fresh tea leaves can be used to process high quality tea (Pang et al., 2022). The quality of fresh tea leaves is not only related to its own genetic characteristics, but also closely related to the ecological factors of the origin of fresh tea leaves, such as latitude, altitude, slope, terrain and topography (Pan et al., 2022), and altitude is one of the most important ecological factors (Chen et al., 2010).

Most of China's famous teas are produced in the region of owning mountains and rivers. The so-called "high mountains produce good tea" mainly refers to the meteorological factors at the altitude that are conducive to the formation of high quality tea. Many scholars have carried out studies on the impact of altitude on the quality of tea. Zhang et al. (2022) found that the contents of theanine, glutamine, threonine, serine and arginine in high altitude green tea were significantly higher than those in low altitude green tea ($P < 0.05$) by analyzing the fresh tea leaves at an altitude from 400 to 1100 m in Anqing City. Huang & Han (2019) analyzed and studied the quality composition contents of Huangshan Maofeng tea at different altitudes. With the increasing of altitude, the amino acid content increased significantly in a linear relationship, while the tea polyphenols and phenol ammonia ratio decreased significantly in an exponential relationship. The influence of altitude on the aroma components of Tieguanyin (Jiang et al., 2021), Xinyang Maojian (Wang et al., 2021) and black tea (Chen et al., 2020) was also studied. It was concluded that the altitude was highly positively correlated with the quality of tea aroma, and the aroma of black tea had a light floral and fruity aroma, with excellent quality. It can be seen that there is an extremely close relationship between altitude and tea quality. Only picking fresh tea leaves at a certain altitude can be processed to high quality tea. So, the sales price of tea at high altitude is often several times higher than that of tea at low altitude (Kfoury et al., 2018).

Therefore, under the same picked standard conditions, the quality of high altitude fresh tea leaves is better than that of low altitude fresh tea leaves, and the purchase price of high altitude fresh tea leaves is often higher than that of low altitude fresh tea leaves. Due to the large profit margin, some tea farmers often picked fresh tea leaves from low altitude areas and pretended to be fresh tea leaves from high altitude areas, and sold them to tea processing factory at a higher price. However, it was difficult for fresh tea leaf purchasers to judge fresh tea leaves at different altitudes based on their own work experience and sensory organs, and the judgment results were still subjective and prone to miscalculation. Therefore, it was urgent to establish a scientific and convenient method to distinguish fresh tea leaves at different altitudes.

*Received 25 Aug., 2022*
*Accepted 17 Oct., 2022*
[1] *Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing, China*
[2] *Institute of Fruit and Tea, Hubei Academy of Agricultural Sciences, Wuhan, China*
*Corresponding author: songzhiyu@caas.cn; wwsspp0426@163.com

Near infrared spectroscopy (NIRS) is a fast, accurate, and non-destructive technique that can be used to replace traditional chemical analysis methods. NIR spectroscopy is a powerful analytical tool widely applied in agricultural (Zhou et al., 2015), petrochemical (Wu et al., 2014), textile (Tavanaie et al., 2015), and pharmaceutical industries (Lee et al., 2011). In terms of food analysis, near infrared spectroscopy can be used to detect food protein, fat, soluble solids, fruit substance and monitor food quality (Shi et al., 2018). NIR spectroscopy has been used in the tea industry, for example, to analyze caffeine and free amino acids, and determine the origins of tea varieties. Backward interval partial least squares (bi-PLS) (Wang et al., 2022b) and back propagation-artificial neural network (BP-ANN) algorithms (Xu et al., 2022; Pranoto et al., 2022), and other pattern recognition systems (Shi et al., 2012), such as the support vector machine, have been used to calibrate tea nutrition models. However, there are few reports on near infrared spectroscopy combined with chemometrics to identify fresh tea leaves at different altitudes, so it is necessary to carry out relevant research.

In this study, the fresh tea leaves were picked in the famous historical tea-Enshi Yulu Tea Reserve in Hubei Province as the research objects. Fresh tea leaves at different altitudes were collected, scanned and obtained their near infrared spectra, and then the spectra were pretreated. The characteristic spectral intervals of fresh tea leaves were screened by the backward interval partial least squares method (biPLS) (Nørgaard et al., 2000), and then the principal component analysis (PCA) was conducted on the characteristic spectral data (Shahdoosti & Ghassemian, 2016). Finally, the least squares support vector machine (LS-SVM) (Chen et al., 2022) was used to establish a NIRS discrimination model for fresh tea leaves at different altitudes, and the robustness of the model was tested by prediction set samples, so as to provide a scientific, objective and convenient new method for the discrimination of fresh tea leaves at different altitudes. And it will also lays a solid technical foundation for the fair acquisition of fresh tea leaves.

## 2 Materials and methods

### *2.1 Sample preparation*

A total of 90 fresh tea leaves samples were collected from Enshi Yulu Conservation Area, Enshi City, Hubei Province. The picking standard was one bud and one leaf, and the picking time was from March to May 2017. There are 30 samples for fresh tea leaves at 450 m < altitude ≤ 600 m (the chemical value was set as 1.00 for the first type samples), 600 m < altitude of ≤ 850 m (the chemical value was set as 2.00 for the second type samples), and 850 m < altitude of ≤ 1000 m (the chemical value was set as 3.00 for the third type samples), respectively. The samples were randomly divided into 2 sets according to the rate of 2:1, including 60 samples in the calibration set and 30 samples in the prediction set, to test the robustness of the calibration set model.

### *2.2 Methods*

*Spectra collection*

NIR spectra were obtained in the reflectance mode using a Thermo Antaris II Fourier transform (FT) NIR spectrometer (Nicolet AntarisII, Thermofisher Scientific, U.S.A.) coupled with an InGaAs detector, a quartz halogen lamp, and an integrating sphere accessory. The samples were placed in a sample cup (ø 30 mm) specifically designed for this application. For each sample, the fresh tea leaves (25 g) was placed into the sample cup according to the procedure specified by the manufacturer. The spectral data were obtained from 10000 $cm^{-1}$ to 4000 $cm^{-1}$ at 3.857 $cm^{-1}$ intervals while rotating the sample cup 360° such that the entire sample was analyzed. Duplicates of each sample were scanned three times. The average spectrum of each sample was employed in following analysis.

*Spectral data analysis*

1) The near infrared spectrum of each sample was converted into 1557 pairs of data points and saved in the excel table. TQ Analyst 9.4.45 software, OPUS 7.0 software and Matlab 2012a software were applied to analyze the spectral data.

2) In order to effectively remove a large amount of background information and noise information in the spectrum and improve the signal-to-noise ratio of the spectrum when modeling, spectral free preprocessing (None), standard normal variable (SNV), multiple scatter correction (MSC), first derivative (FD), second derivative (SD) and their combination spectral preprocessing methods were used to denoise the original spectrum (Wang et al., 2022a), and the best spectral preprocessing method was selected.

3) The biPLS method was used to divide all the pretreated spectral data equally into 22-25 spectral subintervals, and the partial least squares model was established with the n-1 remaining spectral subintervals through the method of leaving-one. When the root mean square error of cross validation (RMSECV) was the lowest, the spectral intervals obtained were the selected characteristic spectral subintervals reflecting the altitude of fresh tea leaves.

RMSECV was calculated as follows (Equation 1):

$$\mathrm{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i^{'} - y_i\right)^2}{n}} \quad (1)$$

Where $n$ is the number of samples in the calibration set, $y_i$ is the true value for sample $i$, and $y_i^{'}$ is the theoretical value for sample $i$ predicted from the calibration set.

4) Principal component analysis (PCA) was performed on the obtained characteristic spectral data. With the principal component score as the input value and the chemical value of fresh tea leaves at different altitudes as the output value, the least squares support vector machine (LS-SVM) method was used to establish NIRS calibration set model of fresh tea leaves at different altitudes, and the prediction set samples were used to test the prediction effect of the calibration set model. The results were expressed by the determination coefficient of cross validation ($R_c^2$), determination coefficient of prediction ($R_p^2$), root mean

square error of cross validation (RMSECV), and root mean square error of prediction (RMSEP). Among them, the larger the $R_p^2$ and the lower the RMSEP, the better the model prediction effect was (Wolfgang & Leopold, 2014).

## 3 Results and discussion

### *3.1 Screening of spectral preprocessing methods*

It can be seen from Figure 1 that the spectrum had more absorption peaks in the long wave band (4000-7000 $cm^{-1}$), mainly reflecting the NIRS absorption information of water -OH and various components in fresh tea leaves, while the NIRS information reflecting the altitude of fresh tea leaves was relatively weak, which was likely to be covered by the NIRS absorption peak information of other components, and will greatly affect the prediction effect of models (Chu, 2021). Therefore, before establishing the model, it was necessary to preprocess the spectrum to filter as much noise information as possible and improve the signal-to-noise ratio of the modeled spectrum. In this study, various spectral pretreatment methods were used to pretreat fresh tea leaves at different altitudes, and partial least squares (PLS) was used to establish the NIRS prediction model. The results were shown in Table 1.

It can be seen from Table 1 that among the models' results established after the above nine spectral pretreatment of the original spectrum, the results of NIRS model established with the original spectrum were the worst ($R_c^2$ = 0.384, RMSECV = 1.412); Among the models established after single pretreatment of spectra, the NIRS model established by FD pretreatment method had better results ($R_c^2$ = 0.494, RMSECV = 1.381), but the prediction results were worse than those established by combined pretreatment method; Among the models established after combined pretreatment method, the NIRS model established by MSC+FD combined method had the best prediction results ($R_c^2$ = 0.670, RMSECV = 1.171). This was because the MSC pretreatment method can remove the noise caused by specular reflection and nonuniformity in the near-infrared diffuse reflectance spectrum, and eliminate the baseline and spectral non repeatability of the diffuse reflectance spectrum; the FD preprocessing method can eliminate baseline drift, strengthen spectral band characteristics, overcome spectral band overlap, and remove drift irrelevant to the same wavelength. The combination of these two methods has improved the pretreatment effect. Compared with the model established by the original spectrum, $R_c^2$ increased by 74.5% and RMSECV decreased by 17.07%. It can be seen that it was very necessary to preprocess the original spectrum before establishing the model, which can effectively eliminate some noise information. The best spectral preprocessing method in this study was the MSC + FD combination method, but the prediction results were still poor, which cannot meet the requirements for accurate prediction of fresh tea leaves samples at different altitudes.

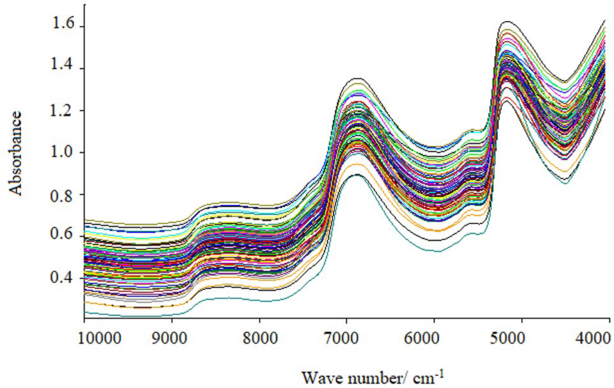


**Figure 1**. Near infrared spectra of fresh tea leaves at different altitudes.

**Table 1**. The results of different pretreatment methods.

| Pretreatment methods | $R_c^2$ | RMSECV |
|---|---|---|
| None | 0.384 | 1.412 |
| SNV | 0.611 | 1.343 |
| MSC | 0.608 | 1.372 |
| FD | 0.494 | 1.381 |
| SD | 0.617 | 1.335 |
| SNV + FD | 0.658 | 1.262 |
| SNV + SD | 0.647 | 1.283 |
| **MSC + FD** | **0.670** | **1.171** |
| MSC + SD | 0.660 | 1.225 |

### *3.2 Screening of characteristic spectral intervals*

Due to the overlapping, covering and cross effects of spectral information, it was necessary to further screen the spectral information intervals closely related to the altitude of fresh tea leaves to improve the prediction accuracy of the model. In this study, the biPLS method was used to screen the characteristic spectral intervals. When all spectral data were divided into 22 spectral subintervals, and RMSECV of the model was the lowest, the spectral subintervals obtained were the filtered characteristic spectral intervals. The results were shown in Table 2.

It can be seen from Table 2 that in the process of establishing biPLS prediction models, when $R_c^2$ was 0.768 and the minimum RMSECV was 0.646, there were 6 spectral subintervals ([4, 6, 9, 12, 19, 22]) modeled, and the corresponding spectral intervals were 4821.2-5091.2 $cm^{-1}$, 5368.9-5638.8 $cm^{-1}$, 6190.4-6460.4 $cm^{-1}$, 7011.9-7281.9 $cm^{-1}$, 8924.9-9191.1 $cm^{-1}$, 9734.9-10000 $cm^{-1}$, respectively. The proportion of the characteristic spectral range in the total spectral range was 27.23%. It can be seen that the biPLS method can screen the characteristic spectral intervals reflecting the fresh tea leaves at different altitudes, which greatly reduced the amount of spectral data and the spectral information to be input into the model, but the prediction accuracy of the model was improved. The $R_c^2$ of the best biPLS model was 14.63% higher than that of the best PLS model, and RMSECV was 44.83% lower than that of the best PLS model.

### *3.3 Principal component analysis*

Before establishing the LS-SVM model, it was required to input as few data as possible, and the principal component analysis of the sample spectra should be carried out first. Therefore, in this study, the principal component analysis was conducted on the characteristic spectral intervals screened by the biPLS method. The cumulative contribution rate of the first five principal components was as follows (Table 3).

It can be seen from Table 3 that after the principal component analysis of the selected characteristic spectral intervals, the contribution rate of the first five principal components decreases rapidly, of which the contribution rate of PC1 was 90.21%, the cumulative contribution rate of PC1-PC5 was 99.99%, and the cumulative contribution rate of the first three principal components was 99.92%. According to the principle of principal component analysis, the information of the first three principal components can represent all the information of the characteristic spectral intervals. The figure of PC1 vs PC2 of three types of fresh tea leaves at different altitudes was as follows.

It can be seen from Figure 2 that the first type of samples (450 < Altitude ≤ 600 m) were mainly distributed in the second and fourth quadrants, the second type of samples (600 < Altitude ≤ 850 m) were mainly distributed in the first and second quadrants, and the third type of samples (850 < Altitude ≤ 1000 m) were mainly distributed in the first and third quadrants. Although the three types of samples showed certain clustering characteristics respectively, there was a serious cross phenomenon between them. Therefore, the LS-SVM method was used to establish a discrimination model for the three types of tea samples.

**Table 2**. Results of characteristic spectral regions selected by biPLS method.

| Spectral numbers | Spectral subintervals | $R_c^2$ | RMSECV |
|---|---|---|---|
| 22 | 22 | 0.705 | 0.781 |
| 21 | 16 | 0.707 | 0.775 |
| 20 | 20 | 0.709 | 0.767 |
| 19 | 7 | 0.711 | 0.759 |
| 18 | 8 | 0.715 | 0.754 |
| 17 | 2 | 0.718 | 0.749 |
| 16 | 13 | 0.720 | 0.746 |
| 15 | 14 | 0.721 | 0.743 |
| 14 | 11 | 0.723 | 0.741 |
| 13 | 12 | 0.725 | 0.738 |
| 12 | 10 | 0.722 | 0.740 |
| 11 | 5 | 0.720 | 0.734 |
| 10 | 3 | 0.728 | 0.732 |
| 9 | 4 | 0.727 | 0.731 |
| 8 | 18 | 0.730 | 0.729 |
| 7 | 19 | 0.731 | 0.727 |
| **6** | **4** | **0.768** | **0.646** |
| 5 | 9 | 0.721 | 0.735 |
| 4 | 12 | 0.698 | 0.845 |
| 3 | 6 | 0.694 | 0.856 |
| 2 | 19 | 0.685 | 0.939 |
| 1 | 22 | 0.674 | 1.049 |

### *3.4 Establishment of LS-SVM model*

In view of the good nonlinear performance of the radial basis function (RBF) kernel function, when using RBF kernel function to establish LS-SVM model (Chen et al., 2022), the key was to select appropriate super parameters $\gamma$ and $\sigma^2$. Super parameter $\gamma$ was used to control model complexity and approximation error; Hyperparameter $\sigma$ (bandwidth coefficient of kernel function) had an important influence on the measurement accuracy of the model. In this paper, the grid search method and 10 equal cross validation method were combined to $\gamma$ and $\sigma^2$ global optimization, and taken the mean square error (MSE) of training set interactive verification as the objective function. The optimization process consists of two steps: coarse selection and fine selection. The number of grid points for coarse selection was 10 × 10. The search step was large, and the error contour was used to establish the optimal parameter range; then, within the parameter range obtained by rough selection, fine selection was carried out with a smaller step length to determine the optimal super parameter combination. The grid point "·" reflected the search range and step size of grid point in the first step, and the curve reflected the error contour and network point "×" reflected the grid point search range and step size in the second step. The parameter optimization result was shown in Figure 3, which the best was $\gamma$ = 387.13, $\sigma^2$ = 0.117.

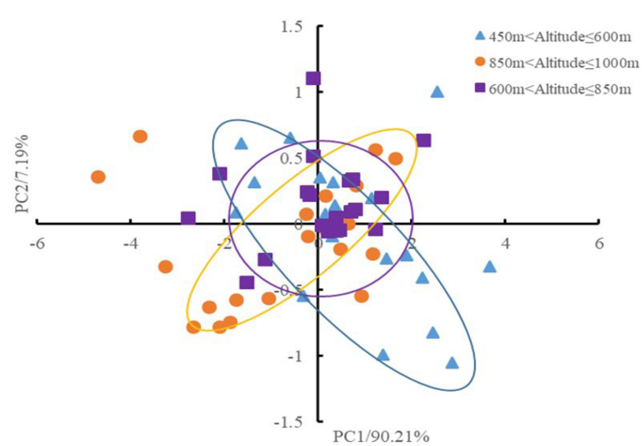


**Figure 2**. PC1 vs PC2 of fresh tea leaves at different altitudes.

**Table 3**. Cumulative contribution rate of the first five principal components.

| Principal components (PC) | PC1 | PC(1-2) | PC(1-3) | PC(1-4) | PC(1-5) |
|---|---|---|---|---|---|
| Cumulative contribution rate/% | 90.21 | 97.42 | 99.92 | 99.97 | 99.99 |

The scores of the first three principal components were as the input values, and the set values of fresh tea leaves at different altitudes were taken as the output values, the LS-SVM method was used to establish NIRS discrimination model for the altitude of fresh tea leaves. The robustness of the calibration set model was tested by using prediction set samples. The results were shown in Figure 4, Figure 5, and Figure 6.

It can be seen from Figure 4 that the true values and the predicted values of the samples in the calibration set model were almost overlap, $R_c^2$ and RMSECV were 0.973 and 0.027, respectively. When 30 samples from the prediction set were used to test the robustness of the calibration set model, $R_p^2$ and RMSEP in the validation set model were 0.968 and 0.034, respectively, and the |prediction deviation |< 0.10 between the true values and the predicted values of the validation set samples indicated that the LS-SVM model had extremely high prediction accuracy, no over fitting phenomenon occurred, and can accurately predict fresh tea leaves at different altitudes. The discrimination rate for both the calibration set and the validation set sample was 100%, and LS-SVM method was applied to realize the quick, accurate and scientific discrimination of fresh tea leaves at different heights. The prediction results of validation set model samples were shown in Table 4.

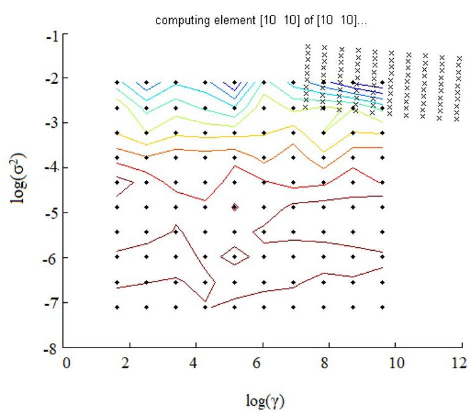


**Figure 3**. Optimation of $\sigma^2$ and $\gamma$ for LS-SVM model.

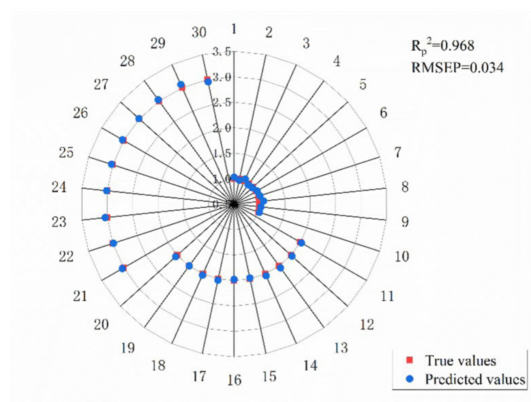


**Figure 5**. The results in prediction set samples by the best LS-SVM model.

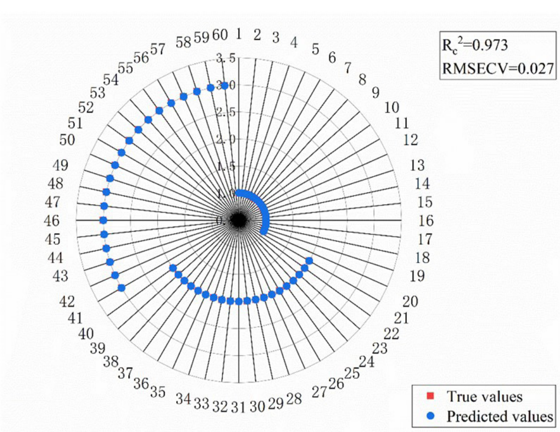


**Figure 4**. The best calibration set model by LS-SVM method.

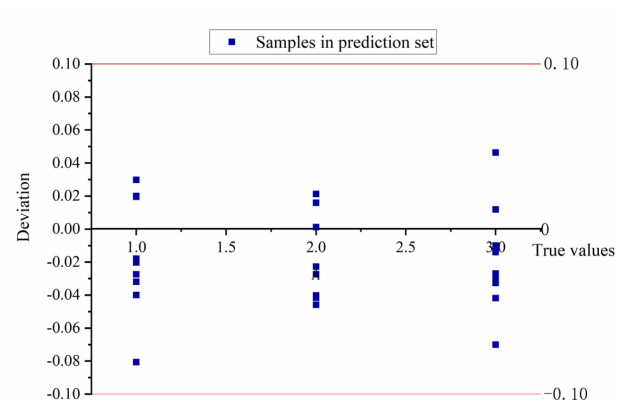


**Figure 6**. The relationship between true values and deviations of prediction set samples.

**Table 4**. The results in prediction set samples by the best LS-SVM model.

| No. | True values | Predicted values | No. | True values | Predicted values | No. | True values | Predicted values |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1.028 | 11 | 2 | 2.023 | 21 | 3 | 3.03 |
| 2 | 1 | 0.98 | 12 | 2 | 2.028 | 22 | 3 | 2.988 |
| 3 | 1 | 1.04 | 13 | 2 | 2.046 | 23 | 3 | 3.042 |
| 4 | 1 | 0.97 | 14 | 2 | 2.04 | 24 | 3 | 3.014 |
| 5 | 1 | 0.98 | 15 | 2 | 1.984 | 25 | 3 | 3.028 |
| 6 | 1 | 1.018 | 16 | 2 | 1.979 | 26 | 3 | 3.027 |
| 7 | 1 | 1.032 | 17 | 2 | 2.028 | 27 | 3 | 3.01 |
| 8 | 1 | 1.081 | 18 | 2 | 2.028 | 28 | 3 | 3.033 |
| 9 | 1 | 1.028 | 19 | 2 | 1.999 | 29 | 3 | 3.071 |
| 10 | 1 | 1.021 | 20 | 2 | 2.042 | 30 | 3 | 2.954 |

## 4 Conclusion

1) The MSC + FD preprocessing method was used to remove some NIRS noise information, and then the biPLS method was used to screen the characteristic spectral intervals reflecting the fresh tea leaves at different altitudes (4821.2-5091.2 $cm^{-1}$, 5368.9-5638.8 $cm^{-1}$, 6190.4-6460.4 $cm^{-1}$, 7011.9-7281.9 $cm^{-1}$, 8924.9-9191.1 $cm^{-1}$, 9734.9-10000 $cm^{-1}$); After the principal component analysis, the scores of the first three principal components were as the input values. When the LS-SVM model was established using the RBF kernel function, and the NIRS model had the best results. The $R_c^2$ and RMSECV were 0.973 and 0.027, respectively.

2) The robustness of the calibration set model was tested by using the prediction set samples, whose $R_p^2$ and RMSEP were 0.968 and 0.034, respectively, indicating the robustness of the calibration set model was good, and there was no over fitting phenomenon. The results of this study has provided a scientific basis for rapid and non-destructive discrimination of fresh tea leaves at different altitudes, and also laid a solid technical support for fair purchase of fresh tea leaves at different altitudes in the future.

## Acknowledgements

This work was supported by the Key R&D projects in Yunnan Province (202102AE090038), Promotion and demonstration of key technologies for fine processing of Sichuan tea (2022YFQ0110), Innovation Project of Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2022-NIAM). Institute-level basic scientific research business expenses project of the Chinese Academy of Agricultural Sciences (S202103-03).

## References

Chen, M., Ni, Y. L., Jin, C. Q., Liu, Z., & Xu, J. S. (2022). Spectral inversion model of the crushing rate of soybean under mechanized harvesting. *Food Science and Technology*, 42, e123221. http://dx.doi.org/10.1590/fst.123221.

Chen, X. H., Sun, H. Y., Qu, D., Yan, F., Jin, W. G., Jiang, H., Chen, C., Zhang, Y. F., Li, C. Y., & Xu, Z. M. (2020). Identification and characterization of key aroma compounds in Chinese high altitude and northernmost black tea (Camellia sinensis) using distillation extraction and sensory analysis methods. *Flavour and Fragrance Journal*, 35(6), 666-673. http://dx.doi.org/10.1002/ffj.3605.

Chen, Y. L., Jiang, Y. M., Duan, J., Shi, J., Xue, S., & Kakuda, Y. (2010). Variation in catechin contents in relation to quality of 'Huang Zhi Xiang' Oolong tea (Camellia sinensis) at various growing altitudes and seasons. *Food Chemistry*, 119(2), 648-652. http://dx.doi.org/10.1016/j.foodchem.2009.07.014.

Chu, X. L. (2021). *Chemometric methods in modern spectral analysis*. Beijing: Chemical Industry Press.

Huang, J. G., & Han, W. Y. (2019). Effect of altitude on the quality of Lushan Yunwu tea. *Chinese Teas*, 4, 19-21.

Jiang, Y., Boorboori, M. R., Xu, Y., & Lin, W. (2021). The appearance of volatile aromas in Tieguanyin tea with different elevations. *Journal of Food Science*, 86(10), 4405-4416. http://dx.doi.org/10.1111/1750-3841.15898. PMid:34494657.

Kfoury, N., Morimoto, J., Kern, A., Scott, E. R., Orians, C. M., Ahmed, S., Griffin, T., Cash, S. B., Stepp, J. R., Xue, D. Y., Long, C. L., & Robbat, A. Jr. (2018). Striking changes in tea metabolites due to elevational effects. *Food Chemistry*, 264, 334-341. http://dx.doi.org/10.1016/j.foodchem.2018.05.040. PMid:29853384.

Lee, M. J., Seo, D. Y., Lee, H. E., Wang, I. C., Kim, W. S., Jeong, M. Y., & Choi, G. J. (2011). In line NIR quantification of film thickness on pharmaceutical pellets during a fluid bed coating process. *International Journal of Pharmaceutics*, 403(1-2), 66-72. http://dx.doi.org/10.1016/j.ijpharm.2010.10.022. PMid:21035529.

Nørgaard, L., Saudland, A., Wagner, J. P., Munck, L., & Engelsen, S. B. (2000). Interval partial least squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy*, 54(3), 413-419. http://dx.doi.org/10.1366/0003702001949500.

Pan, T. H., Yan, R., & Chen, Q. (2022). Geographical origin of green tea identification using LASSO and ANOVA. *Food Science and Technology*, 42, e41922. http://dx.doi.org/10.1590/fst.41922.

Pang, X., Chen, F., Liu, G., Zhang, Q., Ye, J., Lei, W., Jia, X., & He, H. (2022). Comparative analysis on the quality of Wuyi Rougui (Camellia sinensis) tea with different grades. *Food Science and Technology*, 42, e115321. http://dx.doi.org/10.1590/fst.115321.

Pranoto, W. J., Al-Shawi, S. G., Chetthamrongchai, P., Chen, T.-C., Petukhova, E., Nikolaeva, N., Abdelbasset, W. K., Yushchenko, N. A., & Aravindhan, S. (2022). Employing artificial neural networks and fluorescence spectrum for food vegetable oils identification. *Food Science and Technology*, 42, e80921. http://dx.doi.org/10.1590/fst.80921.

Shahdoosti, H. R., & Ghassemian, H. (2016). Combining the spectral PCA and spatial PCA fusion methods by an optimal filter. *Information Fusion*, 27, 150-160. http://dx.doi.org/10.1016/j.inffus.2015.06.006.

Shi, J. Y., Zou, X. B., Zhao, J. W., & Mao, H. P. (2012). Selection of wavelength for strawberry NIR spectroscopy based on BiPLS combined with SAA. *Hongwai Yu Haomibo Xuebao*, 30(5), 458-462. http://dx.doi.org/10.3724/SP.J.1010.2011.00458.

Shi, Q. R., Dai, J., & Jiang, C. Y. (2018). Research progress of near infrared diffuse reflectance technology in food analysis. *Jiangsu Condiment and Subsidiary Food*, 3, 4-8.

Tavanaie, M. A., Esmaeilian, N., & Mojtahedi, M. R. M. (2015). Olive hue visible–near infrared camouflage properties of high speed melt spun poly (ethylene terephthalate) multifilament yarn. *Dyes and Pigments*, 114, 267-272. http://dx.doi.org/10.1016/j.dyepig.2014.11.019.

Wang, H., Cao, X., Yuan, Z., & Guo, G. (2021). Untargeted metabolomics coupled with chemometrics approach for Xinyang Maojian green tea with cultivar, elevation and processing variations. *Food Chemistry*, 352, 129359. http://dx.doi.org/10.1016/j.foodchem.2021.129359. PMid:33735748.

Wang, J. B., Wu, X. H., Zheng, J., & Wu, B. (2022a). Rapid identification of green tea varieties based on FT-NIR spectroscopy and LDA/QR. *Food Science and Technology*, 42, e73022. http://dx.doi.org/10.1590/fst.73022.

Wang, S. P., Liu, P. P., Feng, L., Teng, J., Ye, F., Gui, A. H., Wang, X. P., Zheng, L., Gao, S. W., & Zheng, P. C. (2022b). Rapid determination of tea polyphenols content in Qingzhuan tea based on near infrared spectroscopy in conjunction with three different PLS algorithms. *Food Science and Technology*, 42, e94322.

Wolfgang, H., & Leopold, S. (2014). *Applied multivariate statistical analysis* (2nd ed.). Beijing: Peking University Press.

Wu, C.-F., Wu, T.-G., Hashmonay, R. A., Chang, S.-Y., Wu, Y.-S., Chao, C.-P., Hsu, C.-P., Chase, M. J., & Kagann, R. H. (2014). Measurement of fugitive volatile organic compound emissions from a petrochemical tank farm using open-path Fourier transform infrared spectrometry. *Atmospheric Environment*, 82, 335-342. http://dx.doi.org/10.1016/j.atmosenv.2013.10.036.

Xu, X. J., Ren, S. M., Wang, D. M., Ma, J., Yan, X. W., Guo, Y. L., Liu, X. Q., & Pan, Y. N. (2022). Optimization of extraction of defatted walnut powder by ultrasonic assisted and artificical neural network. *Food Science and Technology*, 42, e53320. http://dx.doi.org/10.1590/fst.53320.

Zhang, Y., Zhu, Y., Lv, H. P., Huang, H., Shao, C. Y., Peng, J. K., & Lin, Z. (2022). Comparative analysis of quality components in baked green tea made from tea plants grown at different altitudes. *Shipin Kexue*, 43(2), 257-268.

Zhou, X., Yang, Z., Haughey, S. A., Galvin-King, P., Han, L., & Elliott, C. T. (2015). Classification the geographical origin of corn distillers dried grains with solubles by near infrared reflectance spectroscopy combined with chemometrics: a feasibility study. *Food Chemistry*, 189, 13-18. http://dx.doi.org/10.1016/j.foodchem.2014.09.104. PMid:26190595.