Original Article
Food Science and Technology

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

DOI: https://doi.org/10.1590/fst.004023

Rapid and non-destructive determination of tea polyphenols content in Chongzhou new loquat tea lines based on near infrared spectroscopy

Ying ZHANG[1,2], Xiaosong LI[1], Hui LI[1], Liang HUANG[1], Jiacheng HUANG[1], Qian TANG[1*]

Abstract

Near infrared spectroscopy (NIRS) combined with multiple algorithms was used to determinate the tea polyphenols content in Chongzhou new loquat tea lines quickly and nondestructively. Samples of 26 Chongzhou new loquat tea lines were collected, then scanning NIRS, pretreating spectral noise information, screening characteristic spectral intervals by backward interval partial least squares, proceeding principal component analysis. Finally, the artificial neural network (BP-ANN) method with three kinds of transfer functions was applied to establish models. The best pretreated method was the combination of standard normal variation (SNV) and first derivative, and the characteristic spectral regions selected were 4381.5-4755.6 cm^{-1}, 4759.5-5133.6 cm^{-1}, 6266.6-6637.8 cm^{-1} and 7389.9-7760.2 cm^{-1}, respectively. The cumulative contribution rate of the first three principal components of the selected characteristic spectra was 95.24%. When the BP-ANN calibration set model was established with the logistic function, NIRS model had the best results, whose root mean square error and determination coefficient of the cross validation were 0.975 and 0.372%, respectively. The root mean square error and the determination coefficient of the prediction set model were 0.962 and 0.400%, respectively. The results showed NIRS can predict the tea polyphenols content in Chongzhou new loquat tea lines quickly and accurately.

Keywords: Chongzhou new loquat tea lines; tea polyphenols; near infrared spectroscopy; backward interval partial least squares; artificial neural network.

Practical Application: Rapid determination of tea polyphenols content in Chongzhou new loquat tea lines.

1 Introduction

Tea [Camellia sinensis (L.) O. Kuntze] originated in Yunnan, Sichuan, Chongqing, Guangxi, Guizhou and other southwest regions of China (Niu et al., 2019). Sichuan is one of the birthplaces of tea trees. Tea germplasm resources are very rich, laying a solid foundation for the cultivation of new varieties of high-quality tea trees with characteristics. Chongzhou loquat tea is a local tea population variety resource in Sichuan formed through long-term natural evolution and semi artificial domestication (Sun et al., 2018). It is characterized by vigorous growth, large loquat like leaves, strong adaptability and cold tolerance, rich genetic background and important breeding potential. It is mainly distributed in the broad-leaved forests in the Longmen mountains in Chongzhou, Sichuan province, 30° 44′ 01”-31° 48′19” north latitude, 103° 03′ 32”-104° 28′ 22” east longitude, 500-1200 m above sea level in the coniferous forest, the habitat belongs to subtropical humid monsoon climate, with annual rainfall of 1300 mm and annual average temperature of 14 °C. According to statistics, wild Chongzhou loquat tea is naturally distributed in Sanlang Town, Wenjingjiang Town, Jiguanshan Town, Huaiyuan Town and other places in Chongzhou, with more than 1200 trees.

26 Chongzhou new loquat tea lines were selected from the population of Chongzhou loquat tea by systematic breeding, which greatly promoted the development and utilization of Chongzhou loquat tea resources. The composition of tea is not only the material basis of good tea quality, but also an important basis for judging the properties of suitable tea. Tea polyphenols, amino acids, caffeine, catechins and other substances are highly related to tea quality. These components and contents are the important basis for the development and utilization of high quality and specific tea plant resources and genetic improvement of varieties. Because the tea polyphenols are especially beneficial to human body, it can decompose fat (Yang et al., 2017), relax stomach (Chen et al., 2010), resist oxidation (Yan et al., 2018), decrease blood sugar (Yamagata et al., 2015) and so on function effect. Therefore, it is very important to study tea polyphenols content in the Chongzhou new loquat tea lines which can be directly developed or used as an excellent parent material for tea breeding.

Tea polyphenols are of great interest owing to their beneficial medicinal properties (Al-Hatim et al., 2022). Increasing evidence shows that polyphenols substances found in tea can enhance general health. Recently, researchers have suggested that antioxidants found in polyphenol substances might play an important role in preventing cardiovascular disease (Nakachi et al., 2000), chronic gastritis (Setiawan et al., 2001), and some cancers (Inoue et al., 2001). Therefore, the tea polyphenol content is an index for

Received 19 Jan., 2023
Accepted 12 Feb., 2023
[1] *Sichuan Agricultural University, Chengdu, PR, China*
[2] *Chongqing Academy of Agricultural Sciences, Chongqing, PR, China*
*Corresponding author: bellazy@126.com

evaluating the Chongzhou new loquat tea lines. Generally, the Folin phenol chemical method (Zhou et al., 2008) is used to determine the tea polyphenols content, but this method is not only time-consuming and laborious, but also pollutes the peripheral environment. Therefore, developing a fast and nondestructive method to determine tea polyphenols content is necessary.

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; Wang et al., 2022a), petrochemical (Wu et al., 2014), textile (Tavanaie et al., 2015), and pharmaceutical industries (Lee et al., 2011). NIRS has been used in the tea industry, for example, to analyze caffeine (Esteban-Diez et al., 2004) and free amino acids, and determine the origins of tea varieties . However, in terms of determining tea polyphenol content in Chongzhou new loquat tea lines, it has not been reported to use NIRS combined with different algorithms to study Chongzhou new loquat tea lines.

In terms of research on Chongzhou new loquat tea lines currently, it mainly focuses on ecological evaluation (Liu et al., 2017), biochemical quality evaluation (Jiang et al., 2020), evaluation of photosynthetic performance of tea plant (Zou et al., 2019), and genetic diversity evaluation (Nyabera et al., 2021) and so on, but there is no relevant research report on the application of near infrared technology to quickly determine the tea polyphenols content in Chongzhou new loquat tea lines. Therefore, the backward interval partial least squares (bi-PLS), principal component analysis (PCA) and back propagation-artificial neural network (BP-ANN) (Xu et al., 2022; Pranoto et al., 2022) were used to establish NIRS models for determination of tea polyphenols content in Chongzhou new loquat tea lines, and the robustness of the model was evaluated by the root mean square error of prediction (RMSEP) and the determination coefficient (Rp^2) in prediction set, so as to provide theoretical basis and scientific and technological support for rapid, non-destructive and objective determination of tea polyphenols content in Chongzhou new loquat tea lines.

2 Materials and methods

2.1 Sample preparation

The total number of 104 Chongzhou new loquat tea lines samples (26 Chongzhou new loquat tea lines, each with 4 samples), processed between March 2018 and September 2018, were obtained from Sichuan Ya'an National Tea Variety Breeding Farm. The sample standard was 1 bud and the first 2 leaves. After the fresh tea leaves were picked, they were microwave sterilized for 2 minutes, then dried at 90 °C, and finally stored in the refrigerator at 4 °C for standby. Before determining tea polyphenols content, all samples were crushed over 80 mesh sieves. Then, the samples were randomly divided into two subsets according to 3:1 ratio, meaning 78 samples in calibration set and 26 samples in prediction set, which were used to establish the calibration models and prediction models, respectively.

2.2 Determination of tea polyphenols

The content of tea polyphenols was determined by the standard Folin phenol method (Zhou et al., 2008). After Chongzhou new loquat tea lines samples grinded, tea polyphenols were distilled under the conditions of 70% methanol and 70 °C water bath. The -OH groups of tea polyphenols were oxidized by Folin phenol reagent, and showed blue color, then the absorbance was measured by a spectrophotometer at the wavelength of 765 nm. The standard curve was drawn by using gallic acid as the standard correction substance used to determine the content of tea polyphenols.

2.3 Spectra collection

NIR spectra were obtained in the reflectance mode using a Thermo Antaris II Fourier transform NIR spectrometer (Nicolet Antaris II, 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, 10 g of the Chongzhou new loquat tea lines samples were placed into the sample cup according to the procedure specified by the manufacturer. The spectral data were obtained from 10,000 cm^{-1} to 4,000 cm^{-1} at 3.857 cm^{-1} intervals while the sample cup rotating 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 (Figure 1).

2.4 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. 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 spectra and improve the signal-to-noise ratio 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., 2022b), and the best spectral preprocessing method was selected;

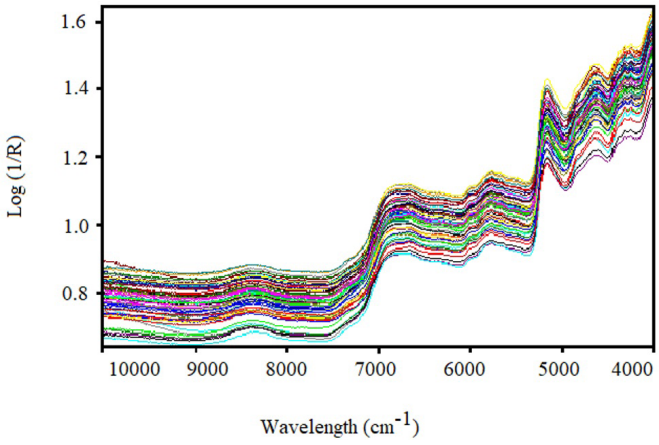

Figure 1. Near infrared spectroscopy of Chongzhou new loquat tea lines samples.

3)The biPLS method was used to divide all the pretreated spectral data equally into 10-22 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 tea polyphenols content in Chongzhou new loquat tea lines samples.

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) PCA and BP-ANN algorithms

PCA (Ghaziri & Qannari, 2015) was performed on the best spectral intervals obtained by biPLS method. BP-ANN algorithm (Liu et al., 2010) was used to establish the NIRS models with the number of principal components (PCs) as the input values and the tea polyphenols content in Chongzhou new loquat tea lines samples as the output values. The results were expressed as 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). A higher R^2 and a lower RMSEP indicated the better prediction effect of the calibration model.

RMSEP was calculated as follows (Equation 2):

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

where n is the number of samples in the prediction set, y_i is the true value of sample i and $y_i^{'}$ is the predicted value of sample i in the prediction set.

R^2 was calculated as follows (Equation 3):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^{'} - y_i)^2}{\sum_{i=1}^{n}(y_i^{i} - \bar{y})^2} \quad (3)$$

where y_i and $y_i^{'}$ are the true value and predicted value of sample i, respectively, and $\bar{y}$ is the average true value of all samples.

3 Results and discussion

3.1 The content of tea polyphenols

The content of tea polyphenols was determined by using the standard method, and the results were shown in Table 1.

As seen from Table 1, the values of the range, mean and S. D. in calibration set were from 21.57% to 34.74%, 30.59% and 3.32, respectively; the values of the range, mean, and S.D. in prediction set were from 21.63% to 31.57%, 26.43% and 2.53, respectively. Thus, it was evident that the range of tea polyphenols content in prediction set was included in the range of samples in calibration set, which laid the foundation for determination of tea polyphenols content more accurately.

3.2 Comparison of pre-processing methods for spectral data

The NIR spectral data of Chongzhou new loquat tea lines samples were pre-treated by various spectrum pre-processing methods. The models of tea polyphenols content were established and the results were shown in Table 2.

Table 2 showed the comparison of 9 model results of tea polyphenols content. The different pre-treatment methods had varying effects on the original spectra of Chongzhou new loquat tea lines. Without pre-processing, the model result was the worst (R_c^2 = 0.543, RMSECV = 1.423%). When different pre-processing methods were used to de-noise the original spectra, the results of models were improved to varying degrees. Compared with the model without pre-processing, the model established by the combined pre-treatment method of (SNV+FD) produced the best values (R_c^2 = 0.711, RMSECV = 1.185%) with RMSECV reduced by 16.7%. Therefore, spectral pre-treatment can effectively improve the signal-to-noise ratio, and this finding was consistent with previous conclusions (Li & Altaner, 2019). However, the results of NIRS models for the tea polyphenols content of Chongzhou new loquat tea lines samples

(Table 2) were still poor. Accurately predicting the tea polyphenols content was still difficult. Therefore, further screening the characteristic spectra intervals that reflecting tea polyphenols content in Chongzhou new loquat tea lines samples was necessary to improve the prediction effect of the model.

Table 1. Classification and tea polyphenols contents of all samples.

Set	Range (%)	Mean (%)	S. D.
Calibration set	21.57-34.74	30.59	3.32
Prediction set	21.63-31.57	26.43	2.53

Table 2. The model results of tea polyphenols content using different preprocessing methods.

Pretreatment methods	R_c^2	RMSECV(%)
None	0.543	1.423
SNV	0.597	1.325
FD	0.614	1.306
SD	0.635	1.245
MSC	0.645	1.208
SNV+FD	**0.711**	**1.185**
SNV+SD	0.638	1.217
MSC+FD	0.642	1.208
MSC+SD	0.695	1.194

SNV: standard normal variation; FD: first derivative; SD: second derivative; MSC: multiple scatter correction.

3.3 Screening of characteristic spectral intervals by biPLS method

BiPLS was used to establish the prediction models with two, three or four spectral intervals. When the RMSECV was the lowest, the modelled spectral intervals were those that exactly reflecting the tea polyphenols content in Chongzhou new loquat tea lines samples. The results were shown in Table 3.

It can be seen from Table 3 that in the process of establishing biPLS models, when R_c^2 was 0.887 and the lowest RMSECV was 0.884%, there were four spectral subintervals ([2, 3, 7, 10]) modeled, and the corresponding spectral intervals were 4377.6 cm^{-1}-4751.7 cm^{-1}, 4755.6 cm^{-1}-5129.7 cm^{-1}, 6262.7 cm^{-1}-6633.9 cm^{-1} and 7386 cm^{-1}-7756.3 cm^{-1}, respectively. The proportion of the characteristic spectral range in the total spectral range was 25.00%. It can be seen that the biPLS method can screen the characteristic spectral intervals reflecting the tea polyphenols in Chongzhou new loquat tea lines samples, which greatly reduced the amount of spectral data and the spectral information to be input into the model, and the prediction accuracy of the model was improved. The R_c^2 of the best biPLS model was 24.75% higher than that of the best PLS model, and RMSECV was 25.40% lower than that of the best PLS model. So, it can be seen that by further screening the characteristic spectral intervals, not only the prediction effect of the tea polyphenols content of Chongzhou new loquat tea lines samples models was further improved, but also the amount of spectral data for modeling was reduced as much as possible. This advantage can reduce the complexity of the model and improve the robustness of the model, which laid a solid foundation for the next step of establishing the artificial neural network model of tea polyphenols content in Chongzhou new loquat tea lines samples.

3.4 Results of PCA analysis

Before establishing the back propagation-artificial neural network (BP-ANN) model, it was required to input as few data as possible, and the principal component analysis of the sample spectra should be carried out firstly. 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 4).

It can be seen from Table 4 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 87.35%, the cumulative contribution rate of PC1-PC5 was 98.56%, and the cumulative contribution rate of the first three principal components was 95.24%. 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. PC1 vs PC2 of Chongzhou new loquat tea lines samples was as follows (Figure 2).

It can be seen from Figure 2 that through principal component analysis of the selected characteristic spectral intervals by biPLS method, the Chongzhou new loquat tea lines samples had certain clustering characteristics. However, most of the samples were scattered and the space distance was far. This may be because the biPLS method was a linear method with limited prediction effect for the tea polyphenols content. Therefore, the next step was to try to apply nonlinear artificial neural network method to further predict the tea polyphenols content in Chongzhou new loquat tea lines samples.

Table 3. Characteristic spectral regions selected by biPLS method.

Spectral numbers	Spectral subintervals	R_c^2	RMSECV (%)
16	13	0.688	1.297
15	14	0.690	1.293
14	11	0.695	1.284
13	12	0.701	1.273
12	16	0.707	1.254
11	5	0.713	1.180
10	1	0.714	1.175
9	4	0.715	1.172
8	6	0.716	1.170
7	15	0.719	1.162
6	8	0.724	1.135
5	9	0.735	1.079
4	**10**	**0.887**	**0.884**
3	7	0.712	1.184
2	3	0.673	1.328
1	2	0.654	1.384

BiPLS: backward interval partial least squares; RMSECV: root mean square error of cross validation.

Table 4. 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/%	87.35	93.25	95.24	97.36	98.56

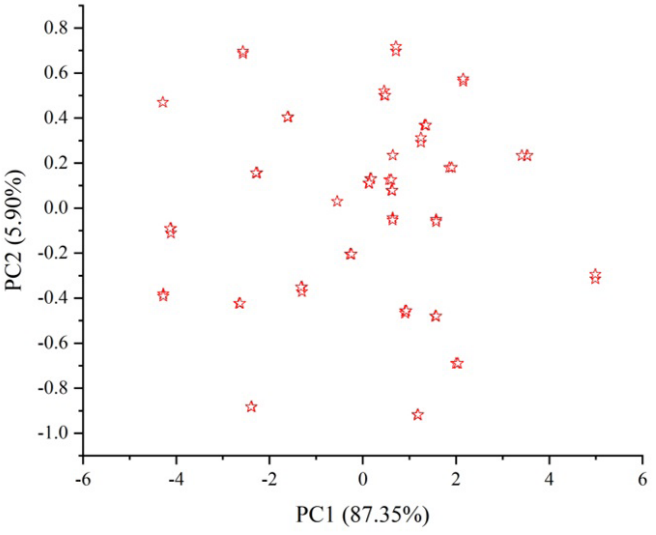

Figure 2. PC1 vs PC2 distribution of Chongzhou new loquat tea lines samples.

3.5 Establishment model of tea polyphenols content by BP-ANN algorithm

The tea polyphenols content model was established by using BP-ANN algorithm and optimised by regulating the number of hidden neurons in the neural network. After multiple tests, the optimal prediction model was calibrated using three PCs input neurons, five hidden neurons, and one output neuron (tea polyphenols content values). During model establishment, varying transfer functions were applied between the transfer layers, and the prediction effect of the model was greatly altered. Three kinds of information transfer functions, namely, linear [-1,1] function, logistic function and tanh function were used in building the BP-ANN models. The model results of these three kinds of transfer functions were compared and shown in Table 5.

It can be seen from Table 5 that the prediction results of the three kinds of transfer function artificial neural network models were different. The relatively poor one was the model established by using the linear [-1,1] function (R_p^2 = 0.891, RMSEP = 0.833%). This may be because Chongzhou new loquat tea lines samples has produced a large number of chemical inclusion components, and the internal chemical components were very complex, so the near infrared spectral information of the components obtained was also extremely complex, which the nonlinear characteristics of spectral information were obvious. Therefore, the prediction results of the artificial neural network model of the tea polyphenols content in Chongzhou new loquat tea lines samples established by linear transfer function [-1,1] were the worst. However, logistic function and tanh function had strong nonlinear characteristics. The prediction results of artificial neural network models established by using these two transfer functions were better than those of linear [-1,1] function models. Among them, the best prediction accuracy was the artificial neural network model of tea polyphenols content in Chongzhou new loquat tea lines samples established by using logistic transfer function (R_p^2 = 0.962, RMSEP = 0.400%) (Lv, 2006). The established artificial neural network model can completely determinate the tea polyphenol content in prediction set samples, and the results of the prediction set model were shown in Figure 3.

It can be seen from Figure 3 that when 26 samples in the prediction set were used to test the robustness of the calibration set model, the true values were almost the same as the predicted values, and the predicted values indicating that the BP-ANN model had extremely high prediction accuracy with logistic function, no over fitting phenomenon occurred, and can accurately predict the tea polyphenols content in Chongzhou new loquat tea lines samples. Figure 3 showed that the best BP-ANN model can accurately predict the tea polyphenols content in Chongzhou new loquat tea lines samples (R_p^2 = 0.962, RMSEP = 0.400%). The results were close to the calibration set model (R_c^2 = 0.975, RMSECV = 0.372%), indicating that the BP-ANN model established by using the logistic transfer function can accurately predict the tea polyphenols content in Chongzhou new loquat tea lines samples. The results in prediction set model samples were shown in Table 6.

Table 5. Results of BP-ANN model with three kinds of transfer functions.

Transfer functions	Calibration set		Prediction set	
	R_c^2	RMSECV(%)	R_p^2	RMSEP(%)
Linear [-1,1]	0.907	0.611	0.891	0.833
logistic	0.975	0.372	0.962	0.400
tanh	0.963	0.406	0.951	0.453

R_c^2: the coefficient of determination for calibration; R_p^2: the coefficient of determination for prediction; RMSEP: root mean square error of prediction; RMSECV: root mean square error of cross validation.

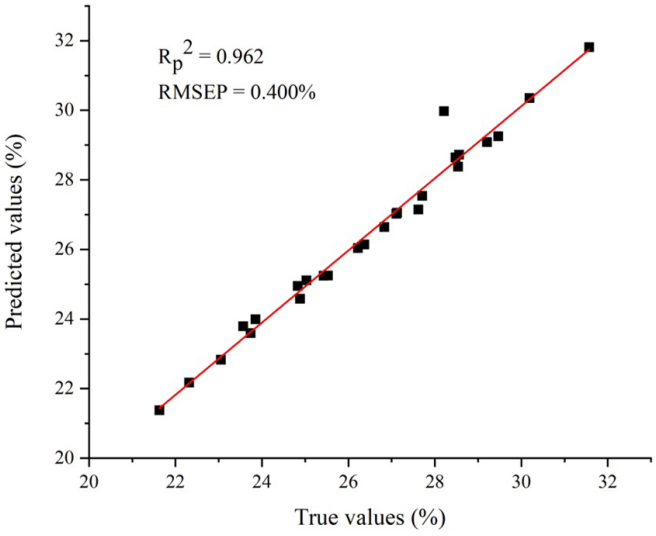

Figure 3. The true values vs predicted values in prediction set samples by the best BP-ANN model. R_p^2: the determination coefficient of prediction set samples; RMSEP: root mean square error of prediction set samples; BP-ANN: back propagation-artificial neural network.

Table 6. The results in prediction set samples by the best BP-ANN model (%).

No.	True values	Predicted values	No.	True values	Predicted values	No.	True values	Predicted values
1	21.631	21.381	10	29.211	29.080	19	28.482	28.640
2	27.711	27.541	11	27.622	27.152	20	25.031	25.112
3	29.471	29.252	12	22.320	22.171	21	30.191	30.352
4	26.223	26.043	13	24.881	24.581	22	28.212	29.971
5	25.530	25.252	14	25.432	25.242	23	23.051	22.831
6	23.742	23.591	15	28.540	28.383	24	24.831	24.951
7	26.831	26.640	16	26.372	26.142	25	23.572	23.792
8	31.572	31.811	17	27.112	27.022	26	23.852	23.991
9	28.561	28.721	18	27.131	27.051			

4 Conclusion

The current developmental trend is to evaluate tea polyphenols content in Chongzhou new loquat tea lines samples non-destructively. In this paper, a robust prediction model (R_p^2 = 0.962, RMSEP = 0.400%) for tea polyphenols content was established by combining NIRS, biPLS, PCA and BP-ANN with the logistic transfer function. The (SNV+FD) was the best preprocessing method to remove NIRS noise information, and then the biPLS method was used to screen the characteristic spectral intervals reflecting the tea polyphenols content in Chongzhou new loquat tea lines samples. The selected characteristic spectral intervals (4377.6 cm^{-1}-4751.7 cm^{-1}, 4755.6 cm^{-1}-5129.7 cm^{-1}, 6262.7 cm^{-1}-6633.9 cm^{-1} and 7386 cm^{-1}-7756.3 cm^{-1}) have eliminated a large amount of irrelevant spectral information. After principal component analysis, cumulative contribution rate of the first three PCs was 95.24%, which were as the input values. When the BP-ANN model was established with the logistic function, and the NIRS model had the best results, whose R_c^2 and RMSECV were 0.975 and 0.372%, respectively. In future applications, these intervals can be used to develop a targeted near-infrared spectral instrument that detects the tea polyphenols content in Chongzhou new loquat tea lines samples without using the full wavelength near-infrared spectral detector. This method can greatly reduce the research cost and advance the commercialisation of the instrument. Additionally, Chongzhou new loquat tea lines samples picked in different years should be collected to enhance the prediction accuracy of the model.

Funding

This study was support by National key research and development program (2022YFD1601401).

References

Al-Hatim, R. R., Al-Alnabi, D. I. B., Al-Younis, Z. K., Al-Shawi, S. G., Singh, K., Abdelbasset, W. K., & Mustafa, Y. F. (2022). Extraction of tea polyphenols based on orthogonal test method and its application in food preservation. *Food Science and Technology (Campinas)*, 42, e70321. http://dx.doi.org/10.1590/fst.70321.

Chen, Y. Q., Zhang, W., Ni, D. J., Cheng, Q., Liu, X. H., & Gan, D. P. (2010). Study on the hypolipidemic effect and antioxidative activity of Hubei Qingzhuan Tea. *Chaye Kexue*, 30(2), 124-128.

Esteban-Diez, I., Gonzalez-saiz, J. M., & Pizarro, C. (2004). An evaluation of orthogonal signal correction methods for the characterisation of arabica and robusta coffee varieties by NIRS. *Analytica Chimica Acta*, 514(1), 57-67. http://dx.doi.org/10.1016/j.aca.2004.03.022.

Ghaziri, A. E., & Qannari, E. M. (2015). Effects of variable selection and processing of NIR and ATR-IR spectra on the prediction of extractive content in Eucalyptus bosistoana heartwood. *Chemometrics and Intelligent Laboratory Systems*, 148, 95-105.

Inoue, M., Tajima, K., Mizutani, M., Iwata, H., Iwase, T., Miura, S., Hirose, K., Hamajima, N., & Tominaga, S. (2001). Regular consumption of green tea and the risk of breast cancer recurrence: follow-up study from the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC), Japan. *Cancer Letters*, 167(2), 175-182. http://dx.doi.org/10.1016/S0304-3835(01)00486-4. PMid:11369139.

Jiang, H., Zhang, M. T., Wang, D. X., Yu, F., Zhang, N., Song, C. K., & Granato, D. (2020). Analytical strategy coupled to chemometrics to differentiate Camellia sinensis tea types based on phenolic composition, alkaloids, and amino acids. *Journal of Food Science*, 85(10), 3253-3263. http://dx.doi.org/10.1111/1750-3841.15390. PMid:32856300.

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.

Li, Y., & Altaner, C. M. (2019). Effects of variable selection and processing of NIR and ATR-IR spectra on the prediction of extractive content in Eucalyptus bosistoana heartwood. *Spectrochimica Acta. Part A: Molecular and Biomolecular Spectroscopy*, 213, 111-117. http://dx.doi.org/10.1016/j.saa.2019.01.060. PMid:30684879.

Liu, F. Q., Li, Q., Rong, Y. T., Wang, L. Y., Xu, L. Y., Yu, F. L., Duan, H. R., & Cheng, H. (2017). Phenotypic diversity of ancient tea tree germplasm resources in Shuangjiang County, Yunnan Province. *Chinese Teas*, 39(4), 22-25.

Liu, Y. D., Sun, X. D., & Ouyang, A. G. (2010). Nondestructive measurement of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLSR and PCA-BPNN. *Lebensmittel-Wissenschaft + Technologie*, 43(4), 602-607. http://dx.doi.org/10.1016/j.lwt.2009.10.008.

Lv, W. Z. (2006). *Modern near infrared spectroscopy analysis technology*. Beijing: China Petrochemical Press.

Nakachi, K., Matsuyama, S., Miyake, S., Suganuma, M., & Imai, K. (2000). Preventive effects of drinking green tea on cancer and cardiovascular disease: epidemiological evidence for multiple targeting prevention. *BioFactors (Oxford, England)*, 13(1-4), 49-54. http://dx.doi.org/10.1002/biof.5520130109. PMid:11237198.

Niu, S. Z., Song, Q. F., Koiwa, H., Qiao, D. H., Zhao, D. G., Chen, Z. W., Liu, X., & Wen, X. P. (2019). Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (Camellia sinensis) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. *BMC Plant Biology*, 19(1), 328. http://dx.doi.org/10.1186/s12870-019-1917-5. PMid:31337341.

Nyabera, L. A., Nzuki, I. W., Runo, S. M., & Amwayi, P. W. (2021). Assessment of genetic diversity of pumpkins (Cucurbita spp.) from western Kenya using SSR molecular markers. *Molecular Biology Reports*, 48(3), 2253-2260. http://dx.doi.org/10.1007/s11033-021-06245-9. PMid:33759053.

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 (Campinas)*, 42, e80921. http://dx.doi.org/10.1590/fst.80921.

Setiawan, V. W., Zhang, Z. F., Yu, G. P., Lu, Q. Y., Li, Y. L., Lu, M. L., Wang, M. R., Guo, C. H., Yu, S. Z., Kurtz, R. C., & Hsieh, C. C. (2001). Protective effect of green tea on the risks of chronic gastritis and stomach cancer. *International Journal of Cancer*, 92(4), 600-604. http://dx.doi.org/10.1002/ijc.1231. PMid:11304697.

Sun, Q., Lan, X. G., Zhang, Y. Q., & Yu, F. L. (2018). West sichuan tea tree. *Chinese Teas*, 40(10), 30-34.

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, 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 (Campinas)*, 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 (Campinas)*, 42, e94322. http://dx.doi.org/10.1590/fst.94322.

Wu, C.-F., Wu, T., 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 artificial neural network. *Food Science and Technology (Campinas)*, 42, e53320. http://dx.doi.org/10.1590/fst.53320.

Yamagata, K., Tagami, M., & Yamori, Y. (2015). Dietary polyphenols regulate endothelial function and prevent cardiovascular disease. *Nutrition (Burbank, Los Angeles County, Calif.)*, 31(1), 28-37. http://dx.doi.org/10.1016/j.nut.2014.04.011. PMid:25466651.

Yan, S. S., Shao, H. J., Zhou, Z. H., Wang, Q., Zhao, L., & Yang, X. (2018). Non-extractable polyphenols of green tea and their antioxidant, anti-α-glucosidase capacity, and release during in vitro digestion. *Journal of Functional Foods*, 42, 129-136. http://dx.doi.org/10.1016/j.jff.2018.01.006.

Yang, X., Huang, M., Qin, C., Lv, B., Mao, Q., & Liu, Z. (2017). Structural characterization and evaluation of the antioxidant activities of polysaccharides extracted from Qingzhuan brick tea. *International Journal of Biological Macromolecules*, 101, 768-775. http://dx.doi.org/10.1016/j.ijbiomac.2017.03.189. PMid:28373046.

Zhou, W. L., Xu, J. F., & Xu, L. (2008). *Determination of total polyphenols and catechins content in tea*. Beijing: China Light Industry Press.

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.

Zou, Y., Xu, Y., Chen, S. X., Tan, L. Q., & Tang, Q. (2019). Evaluation of photosynthetic performance of 25 tea varieties based on factor and cluster analysis. *Journal of Yunnan Agricultural University*, 34(1), 89-96.