

# Detection of Adulteration in Camellia Oil Using Near-Infrared Spectroscopy

Qingsong Luo, Yaru Yu, Qiang Xu, Yang Chen, Xiao Zheng\*

School of Mechanical Engineering, Wuhan Polytechnic University

**Abstract:** Near-infrared spectroscopy (NIRS) combined with chemometrics analysis was used in this study to qualitatively and quantitatively determine the adulterated *Camellia* oil. A binary model was constructed for determining both the authenticity and the number of adulterated contents. NIRS combined with support vector machine classification was used to establish a full spectral model and a selected spectral model via competitive adaptive heavy-weighted sampling and backward interval partial least squares. Notably, both of them were proved to be suitable for determining the authenticity of *Camellia* oil. NIRS combined with support vector machine regression may be used to predict the amount of adulterated content in *Camellia* oil because of the high model correlation coefficient ( $R$  was higher than 99%, and the maximum mean square error was 0.0605).

## 1 INTRODUCTION

*Camellia* oil is a natural high-grade woody oil with an extremely high nutritional value<sup>[1]</sup>. It is one of the edible vegetable oils recommended by the "Outline of China's Food Structure Reform and Development Plan.". The Food and Agriculture Organization of the United Nations also listed it as one of the best healthcare plant oils<sup>[2]</sup>. However, various counterfeit and adulterated oils have flooded into the market due to the high market price and handsome profit related to *Camellia* oil. At present, the identification of adulterated plant edible oil is based mainly on compositional analysis with chromatography, mass spectrometry, spectroscopy, and nuclear magnetic resonance<sup>[3]</sup>. These techniques have several common problems: they are usually expensive, time-consuming, and difficult to operate. Therefore, an inexpensive, efficient, and convenient detection method to realize the effective eradication of *Camellia* oil adulteration is urgently needed<sup>[4]</sup>. The near-infrared spectroscopy (NIRS) technique has rapidly developed in China in recent years. It has been successfully employed for other rapid detection applications<sup>[5-7]</sup>. A rapid detection method suitable for determining adulterated *Camellia* oil was established based on the aforementioned technique.

## 2 EXPERIMENTAL MATERIALS AND EQUIPMENT

### 2.1 Materials and preparation methods

A variety of domestic and international brands of *Camellia* oil, soybean oil, and rapeseed oil were purchased online or from a local supermarket to achieve sample diversity. Tea seeds of various origins were also purchased and used to prepare fresh oil using a laboratory oil press. The blended oil samples were configured as the following percentage of mass: if the blending ratio was 35% or less, the concentration gradient was 2%; if it was more than 35%, the concentration gradient was 5%. A total of 256 oil test samples were used. Of these, 240 samples were adulterated samples, 8 of *Camellia* oil (including the freshly prepared ones), 4 of soybean oil, and 4 of rapeseed oil. A summary of the experimental samples is shown in Table 1.

**Table 1.** Details of the experimental samples

Type	Camellia oil	Camellia oil blended with soybean oil	Soybean oil	Camellia oil blended with rapeseed oil	Rapeseed oil
Number	8	120	4	120	4
ID	1	2	3	4	5

### 2.2 Experimental instruments and equipment

The experimental instruments mainly included the edible

oil quality rapid detector based on laser NIRS, the "Delong" household mini oil press, the TG16-II medical centrifuge, and the JA1003N electronic balance<sup>[7]</sup>. The basic laboratory equipment used in the experiment

mainly included pipettes, sharp-mouthed centrifuge tubes, and 2-mm cuvettes<sup>[8]</sup>.

### 2.3 Sample spectrum acquisition

The room temperature was maintained at 26°C, and the centrifuge tubes containing the experimental sample were heated to 60°C in a water bath. Subsequently, the samples were transferred to a cuvette using a pipette, filling about two-thirds of the full volume, following which the sample spectral data were collected using a spectrophotometer. The absorbance value of each sample was measured three times, and the average was taken as the final absorbance value. The raw spectral data of the experimental samples is shown in Figure 1.

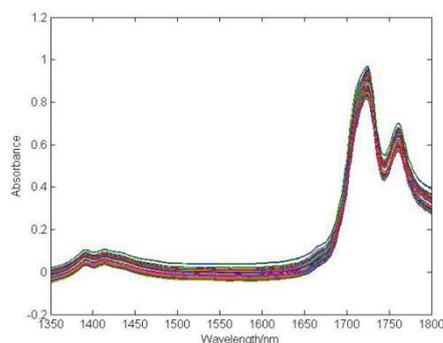


Figure 1. Raw spectra of the experimental samples.

## 3 RESULTS AND DISCUSSION

### 3.1 Spectral data processing method

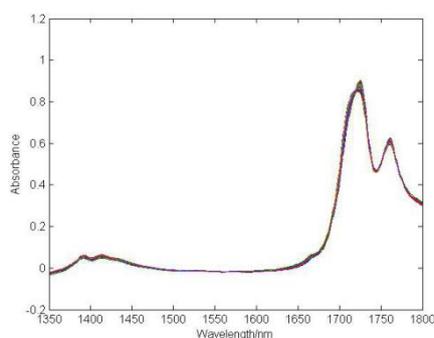


Figure 2. NIRS of Camellia oil samples after SNV-DT.

MATLAB was used to process the spectral data so as to minimize the unfavorable interference, which typically leads to spectral overlap, low signal-to-noise ratio, and nonspecific spectrum<sup>[9]</sup>. The preprocessing step involved two methods, multi-scatter correction (MSC) and de-trending technique (DT). Then, partial least squares, backward interval partial least squares (BiPLS), and competitive adaptive weighted sampling (CARS) were further applied to extract the characteristic wavelengths<sup>[10]</sup>.

In spectral modeling, qualitative identification was based on support vector machine classification (SVC); for quantitative prediction, support vector machine regression (SVR) was used, and the loss function was constructed. A kernel function was employed to establish the connection because both calculations had linear and nonlinear regressions. The choice of penalty parameter  $C$  and the kernel function parameter  $g$  (for the radial basis function kernel) had a great influence on the accuracy of the model when support vector machine (SVM) modeling was used. Therefore, cross-validation (CV) and genetic algorithm were used for preliminary screening in this study to choose the optimal  $C$  and  $g$ , and then the optimal parameters ( $C$  and  $g$ ) were selected for modeling<sup>[11]</sup>.

### 3.2 Preprocessing of the spectral data

MSC and standard normal variate (SNV) transformation plus DT were used for the preprocessing of NIRS. The preprocessed spectra of *Camellia* oil samples are shown in Figures 2 and 3.

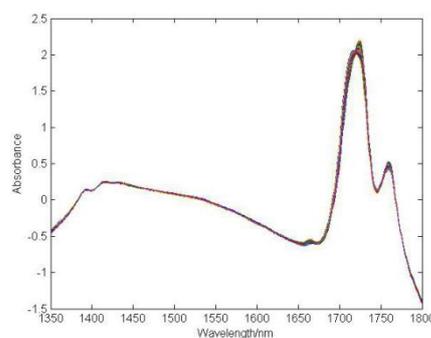


Figure 3. NIRS of Camellia oil MSC.

### 3.3 Qualitative determination of the authenticity

First, the preprocessed spectral data and the spectral data optimized by CARS and BiPLS characteristic variables were separately used as the input data of the SVC model with optimal SVC modeling parameters ( $C$  and  $g$ )

selected by CV. A qualitative SVC model for determining the authenticity of *Camellia* oil was constructed for each input. Table 2 shows the NIRS-SVC model parameters after imputing data with different preprocessing methods and the accuracy of each model.

Table 2. Prediction results of NIRS-SVC models

Oil type	Preprocessing method	Number of variables	Parameters ( $C$ and $g$ )		Accuracy/%	
			$C$	$g$	Calibration set	Prediction set

Camellia oil	MSC	451	1024	16	97.3958	92.1875
	SNV	451	1024	2	98.9583	92.1875
	SNV-DT	451	1024	2	98.4375	92.1875
	MSC-CARS	38	1024	32	87.5	85.9375
	SNV-CARS	48	1024	2	88.5417	90.625
	SNV-DT-CARS	56	512	4	90.625	92.1875
	MSC-BiPLS	134	1024	2	84.375	90.625
	SNV-BiPLS	135	512	16	96.875	93.75
	SNV-DT-BiPLS	90	1024	16	96.3542	92.1875

Table 2 shows that, overall, the qualitative NIRS model constructed with SVC could efficiently and qualitatively determine the authenticity of *Camellia* oil. In particular, input preprocessed by BiPLS resulted in a higher accuracy and faster modeling compared with input preprocessed by CARS. Thus, BiPLS seemed to have a better predictive ability.

Two categories of blended *Camellia* oil were used to construct the quantitative prediction model: first, *Camellia* oil blended with soybean oil (C + S, including 4 pure *Camellia* oil samples, 120 *Camellia* oil + soybean oil samples, and 4 soybean oil samples) and *Camellia* oil blended with rapeseed oil (C + R, including 4 pure *Camellia* oil samples, 120 *Camellia* oil + rapeseed oil samples, and 4 rapeseed oil samples). The raw spectrum data of the sample are shown in Figures 4 and 5.

### 3.4 Quantitative prediction of adulterated content

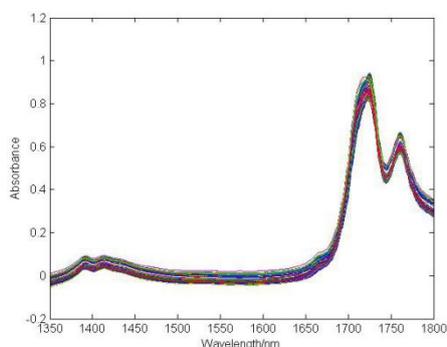


Figure 4. Raw NIRS of C + S samples.

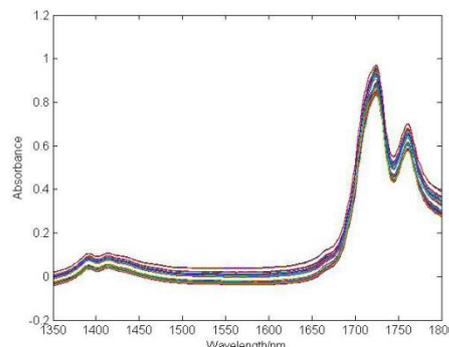


Figure 5. Raw NIRS of C + R samples.

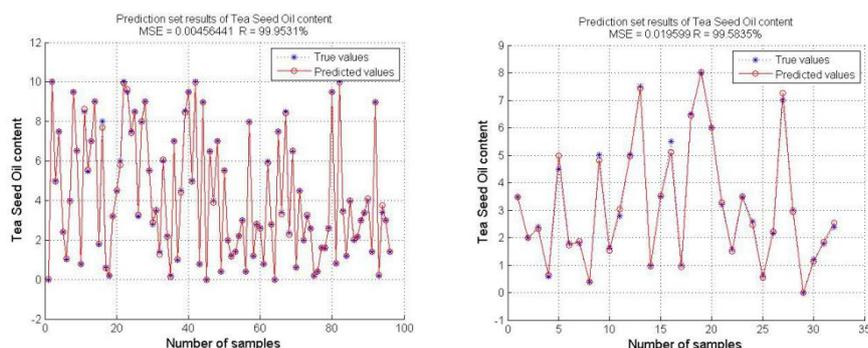
#### 3.4.1 C + S samples

The preprocessed NIRS data of C + S samples, with and without BiPLS variable selection, were used as input to construct the adulterated content prediction SVR model for C + S samples. The VC method was used to find the

optimal parameters (*C* and *g*) of the SVR model. Table 3 shows the prediction results of the SVR models for C + S samples. Figure 6a and 6b show the calibration set results and prediction results of the prediction models based on SNV-BiPLS, respectively.

Table 3. Quantitative prediction of adulterated content in G + C samples

Preprocessing methods of NIRS data	Number of variables	Parameters (C and g)		Calibration set		Prediction set	
		C	g	R (%)	MSE	R (%)	MSE
MSC	451	1024	16	99.987	0.0013	99.5871	0.0195
SNV	451	1024	0.25	99.8683	0.0128	99.4222	0.0307
SNV-DT	451	1024	1	99.9856	0.0014	99.6446	0.0169
MSC-BiPLS	112	256	64	99.8751	0.0121	99.274	0.0337
SNV-BiPLS	315	1024	1	99.9531	0.0046	99.5835	0.0196
SNV-DT-BiPLS	112	1024	1	99.8566	0.0140	99.3326	0.0339



**Figure 6(a).** Calibration set results of the prediction models based on SNV-BiPLS for G + C samples.

**Figure 6(b).** Prediction set results of the prediction models based on SNV-BiPLS for G + C samples.

Table 3 shows that the quantitative prediction model for G + C samples could produce accurate results using NIRS combined with support vector machine regression, as the predicted  $R$  values were all greater than 99%, and the maximum MSE was 0.0337. SNV-BiPLS-SVR (Fig. 6) was the best among the constructed models because the relatively small number of selected variables (315) made the prediction of this model faster and the  $g$  parameter ( $g = 1$ ) indicated that the model had the strong predictive ability. Besides, the  $R$  value of calibration and prediction sets was 99.9531% and 99.5835%, respectively, indicating a high correlation, and the MSE

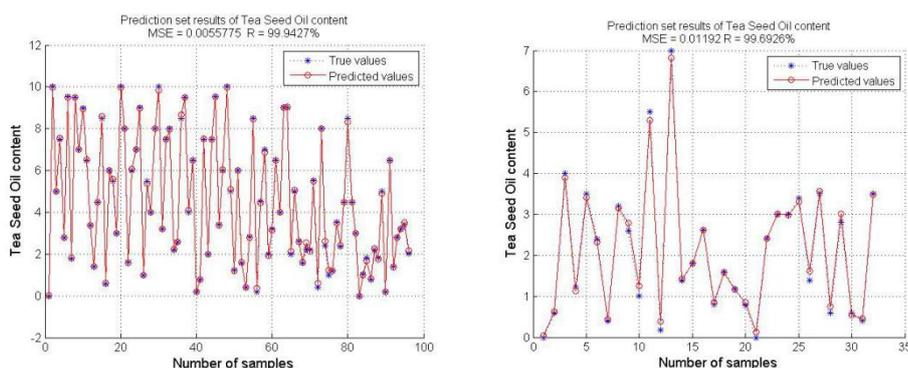
value was 0.0046 and 0.0196, respectively, suggesting that the relative error was minor and the prediction results were reliable.

### 3.4.2 C + R samples

Table 4 shows the prediction of the SVR model for G + C samples using preprocessed data with and without BiPLS characteristic variable selection. Figure 7a and 7B shows the correction set and prediction set results of the SNV-DT-BiPLS-SVR model, respectively.

**Table 4.** Quantitative prediction of adulterated content in G + R samples

Preprocessing methods of NIRS data	Number of variables	Parameters (C and g)		Calibration set		Prediction set	
		C	g	R (%)	MSE	R (%)	MSE
MSC	451	1024	16	99.9941	0.0006	99.7927	0.0063
SNV	451	1024	1	99.9946	0.0005	99.7943	0.0063
SNV-DT	451	1024	1	99.9926	0.0007	99.7269	0.0074
MSC-BiPLS	44	1024	16	99.5762	0.0405	98.9463	0.0326
SNV-BiPLS	44	1024	1	99.5765	0.0404	98.9566	0.0320
SNV-DT-BiPLS	157	256	4	99.9427	0.0056	99.6926	0.0119



**Figure 6(a).** Calibration set results of the prediction models based on SNV-BiPLS for G + R samples.

**Figure 6(b).** Prediction set results of the prediction models based on SNV-BiPLS for G + R samples.

Table 4 shows that the quantitative prediction of adulterated oil content could be accurately realized using NIRS combined with the support vector machine regression. Among these models, the minimum predicted  $R$  was 98.9463% and the maximum MSE was 0.0326. The SNV-DT-BiPLS-SVR model (Fig. 7) with SNV-DT preprocessing and BiPLS-selected 157 characteristic variables was the best model because its modeling speed was fast, and the SVR model parameters ( $C = 256$ ;  $g = 4$ )

rendered it a greater learning ability and thus could be used for practical application. Moreover, the  $R$  value of the correction and prediction sets was 99.9427% and 99.6926%, respectively, indicating a high correlation; and the MSE value was 0.0056 and 0.0119, respectively, suggesting that the relative error was minor and the prediction results were reliable.

## 4 CONCLUSIONS

In this study, NIRS models were constructed for the qualitative and quantitative determination of the adulterated *Camellia* oil (blended with soybean and rapeseed oils). Notably, in the qualitative model, the raw spectra data were preprocessed by MSC/SNV-DT and the selected characteristic variables by CARS and BiPLS were also used. In the quantitative model, in addition to the full spectral data model, the BiPLS-selected characteristic variables were used to construct the characteristic variable model.

The SVC model constructed for *Camellia* oil gave a classification accuracy rate of 99.4975% in the qualitative test, suggesting that it could be used as an accurate method for detecting the adulteration in edible oils. More interestingly, the *R* values were greater than 98% and the MSE values were less than 0.06 for both G + C and G + R samples in the quantitative prediction, indicating that NIRS combined with SVR represented a reliable method for the binary determination of blended edible oils. The gradient of artificial impurity may be further subdivided based on these results in subsequent studies, thereby expanding the reference database for adulterated oils.

## Acknowledgment

Funding: 2016 Grain Science and Technology Innovation and Achievement Transformation Program in Hubei Province (20165104), Wuhan Science and Technology Research Program (2013010501010147).

## REFERENCES

1. Yang Fan, Xue Changyong. Research progress on nutritional characteristics and effects of common edible oils[J]. Food and Nutrition in China, 2013, 19(03): 63-66.
2. Cheng Zhengzai, Gong Kai, Luo Can et al. Edible oil and human health [J]. Chemistry Teaching, 2014 (11): 81-86.
3. Zeng Yanming, Du Shuxin. Research progress of on-line food safety detection based on spectral analysis[J]. Science and Technology of Food Industry, 2010, 31(11): 389-393.
4. Liu Bo, Li Dan. Health care function and application status of tea seed oil [J]. Tea Industry Bulletin, 2008 (03): 112-114.
5. Zhang Yan. Rapid identification and detection of edible oil adulteration by near-infrared spectroscopy [D]. Henan University of Technology, 2015.
6. Ma Wei, Wang Xiu, Li Weiguo. Application of Near Infrared Spectroscopy (NIRs) Detection Technology in Fruit and Vegetable Production in Facilities[J]. Agricultural Engineering Technology, 2015(31): 60-63.
7. Zhu Yutian, Li Jincai, Gao Sujun et al. Research progress of near-infrared spectroscopy in the field of rapid detection of edible oils [J]. China Oils and Fats, 2017, (7): 140-143.
8. Zeng Lulu, Tu Bin, Yin Cheng et al. Qualitative and quantitative analysis of peanut oil blending with laser near-infrared combined with SVM[J]. Journal of the Chinese Cereals and Oils Association, 2016, 31(08): 126-130+137.
9. Tu Bin, Song Zhiqiang, Zheng Xiao et al. Pseudo-qualitative analysis of rice oil based on laser near-infrared[J]. Spectroscopy and Spectral Analysis, 2015.
10. Chen Jinfeng. Research and application of support vector machine regression algorithm [D]. Jiangnan University, 2010.
11. ZHU Fang, ZHAO Qingping, WANG Jiangtao. Study on SVM Classification Application Based on RBF Kernel[J]. Journal of Tangshan University, 2017.