Research Article

Determination of Multicomponents in *Rubi Fructus* by Near-Infrared Spectroscopy Technique

Wenjun Du,[1] **Chunyan Wu** ⬤,[2] **Hesong Yu,**[1] **Qingran Kong,**[2] **Yunjian Xu,**[1] **and Weidong Zhang** ⬤[1]

[1]*Jinhua Municipal Central Hospital, Jinhua 321000, China*
[2]*Taizhou Vocational College of Sciences & Technology, Taizhou 318020, China*

Correspondence should be addressed to Weidong Zhang; phar2008@163.com

Objective. Rubi Fructus (RF) is a dry aggregate fruit of *Rubus* (Rosaceae). It has shown significant pharmacological effects such as antioxidation, hypoglycemic, and anti-inflammatory. A combination of near-infrared (NIR) spectroscopy and partial least squares regression (PLSR) under seven different spectral data preprocessing techniques was used to determine the performance of quantitative analysis correction models which employed moisure, ellagic acid, and total flavonoids as indicators of RF. *Methods.* Ninety-seven different RF batches were collected for NIR spectra. By using primary analysis techniques such as drying method, high-performance liquid chromatography (HPLC), and ultraviolet visible spectrophotometry (UV-Vis), the contents of moisure, ellagic acid, and total flavonoids were determined. The NIR spectral data and the primary analysis method data were correlated through PLSR. Seven methods were used for pretreating the spectral data, including no spectral pretreatment, first derivative, standard normalized variate, multiple scattering corrections, elimination of constant offset, and minimum maximum normalization. The quantitative analysis correction models adopted PLSR chemometrics for moisture, ellagic acid and total flavonoids were developed, and their effectiveness was evaluated using the correlation coefficient (R), ratio of prediction to deviation (RPD), and root mean square error (RMSE). *Results.* The first derivative was combined with variable standardization, elimination of constant offset, and multiple scattering corrections, respectively, to pretreat the PLSR models for moisture, ellagic acid, and total flavonoids. The R-values of the PLSR models for moisture, ellagic acid, and total flavonoids were, respectively, 0.9788, 0.9468, and 0.9748, all of which were higher than 0.90, and the RPD values were 4.9, 3.1, and 4.5, respectively, which were all larger than 3.0. The RMSE ratios of the calibration set and the test set were 0.98, 0.94, and 1.0, respectively. *Conclusion.* The R-values of the NIR-PLSR models for moisture, ellagic acid, and total flavonoids are all greater than 0.90 after suitable pretreatments, indicating that the models are reliable. The RPD values are more than 3.0, which indicate that the models are good and useable for quality control. The RMSE ratios are closed to 1, indicating that the calibration set and test set had same distribution and the models were not overfitting indicating good predictability.

1. Introduction

Rubi Fructus (RF) belongs to the genus *Rubus* of Rosaceae. Its dry aggregate fruit has the effects of tonifying the kidney, consolidating essence, shrinking urine, nourishing the liver, and improving eyesight [1]. RF contains a lot of bioactive substances such polysaccharides, flavonoids, phenols, and terpenoids [2], which have a lot of pharmacological effects such as antioxidant [3], hypoglycemic [4, 5], immune regulation [6], liver cell protection [7], antibacterial [8], and anti-inflammatory [9]. Additionally, when combined with medications, RF has synergistic effects, such as anti-Alzheimer's activity [10] and reducing gout [11]. At present, commonly primary analysis methods for the effective medicinal components in RF mainly include ultrahigh performance liquid chromatography [12], high-performance liquid chromatography [13], liquid chromatography mass spectrometry [14, 15], gas chromatography ion migration spectroscopy [16], and others. The primary analysis methods, regarded as the first level analysis method, are more relatively accurate, but operation procedures are comparatively onerous, workloads are heavy for large-scale

testing, and there is high expense and pollution. As a secondary analysis technique, near-infrared (NIR) spectroscopy analysis technology is quick, pollution-free, and does not require sample preparation, making it more suitable for large-scale and regular rapid detection needs [17]. NIR spectral analysis technology has been developed to this day, mainly extending fields including petroleum and petrochemical, medical clinical, life sciences, pharmaceuticals, food, tobacco, textiles, quality supervision, environmental protection, etc. [18]. In conjunction with appropriate chemometrics, NIR spectral analysis technology can swiftly and simultaneously calculate multiple sample components. This work uses the NIR spectroscopy technology to create a rapid and nondestructive multicomponent analysis approach in order to support the healthy and sustainable development of RF industrialization. RF adulteration and counterfeiting should be prevented to some extent by ensuring high-quality RF products at reasonable pricing. It would have significant applications in drug inspection and market supervision.

2. Materials and Methods

2.1. Materials. RF cultivars were obtained from distinct regions including Pan'an, Jin'yun, Bo'zhou, San'men, Guang'de, Zhe'rong, Yong'jia, Yuan'qiao, Xian'ju, Yue'qing Yan'dang'shan, Rui'an, Fu'shan, Wu'yi, Ning'guo, Yi'wu, Sheng'xian, Qian'dao Lake, Feng'hua, Lin'an, and De'yang. Ellagic acid (97% purity) and Rutin (96% purity) standards were purchased from Baoji Chenguang Biotechnology Co., Ltd. Acetonitrile manufactured by MEKER in Germany was of chromatographic grade. All other organic solvents were of domestic analytical purity.

The MATRIX-Fnear-infrared spectrometer was manufactured by Bruker in Germany. The UV-8000 Ultraviolet-Visible Spectrophotometer was from Shanghai Yuanxi Instrument Co., Ltd. ACQUITY UPLC H-Class PLUS high-performance liquid chromatography system (Waters Corporation, USA) was equipped with PDA detector and Empower software chromatography workstation.

2.2. Component Determinations

2.2.1. Sample Preparation. Every locality of RF cultivars was collected with at least 3 batches. There were total collected ninety-seven batches and numbered from one to ninety-seven, each weighing 250 g. Every batch was pulverized by 80 mesh to obtain auxiliary sample powder for all the subsequent analyses after collection.

NIR spectra of powdered RF were acquired with air as the background. Each spectrum resulted from an average of 32 scans with a resolution of 8 cm^{-1} in the wavenumber interval of $4000 \text{ cm}^{-1} \sim 12000 \text{ cm}^{-1}$. A total of three spectra per sample were collected, and their average spectrum was regarded as the original spectrum.

2.2.2. Moisture Content. The drying method was performed as described in Pharmacopoeia 2020 Edition (the second method of General Rule 0832) [1]. Firstly, milled RF

$(2 \text{ g} \sim 5 \text{ g})$ was tiled with a thickness of no more than 5 mm in a flat bottle which dried to constant weight using in triplicate. Then, the flat bottles loaded with milled RF were placed and dried in an oven (DHG-9055A, Shanghai Yiheng, China) at $100°C \sim 105°C$ for 5 hours. Next, they were moved into a glass desiccator to cool and precisely weighed after 30 minutes. They were dried again at the above temperature for an hour and circulated until the difference between two successive weighing was no more than 5 mg. Calculate the moisture content in the RF sample.

2.2.3. Total Flavonoid Content. The total flavonoid content in RF was detected by the UV spectrophotometer (UV-1800PC-DS2, MAPADA, China) [19]. Rutin standard solutions were used for calibration. The preparation of the rutin standard curve was as follows: an appropriate amount of different concentration rutin solutions was taken and dissolved with 10 mL of ethanol. 1.0 mL of a $NaNO_2$ (5%, m/m) solution was added and laid aside at room temperature ($20°C \pm 5°C$) for 6 minutes. Then, 1.0 mL of an $Al(NO_3)_3$ (10%, m/m) solution was put in and also at room temperature for 6 minutes. Afterward, 10 mL of a NaOH (4%) solution was contributed and diluted to 25 ml with deionized water. After 15 minutes of the reaction, the absorbance was measured at 510 nm. The blank control was with no sample solution.

(1) Preparation of Test Solution. Weigh 1.0 g of milled RF with a thousandth electronic balance, add the appropriate amount of ethanol to dissolve it, and fix the volume to 10 ml. Then, precisely transfer 0.2 ml above solution into a 25 ml volumetric flask, and add 10 ml of ethanol solution. Whereupon referring to the preparation of the standard curve, measure the absorbance value of the test solution at 510 nm, substitute it into the standard curve to calculate the volume concentration of total flavonoids, and substitute it into formula (1) to obtain the mass concentration of total flavonoids in RF. The intrinsic absorption of samples was performed by replacing all the reagents (12 mL) with water. The results of the total flavonoid content were expressed as rutin equivalents (mg rutin/g dry RF), and each extract was analyzed in triplicate.

$$C_m = C_0 \times V \times \frac{F/M_0}{1000}, \qquad (1)$$

C_m: mass concentration of total flavonoids, mg/g; C_0: volume concentration of total flavonoids, ug/mL; V: constant volume, mL; M_0: mass of *Rubi Fructus*, g; F: dilution factor;

2.2.4. Ellagic Acid Content. The ellagic acid content detection was carried out upon high-performance liquid chromatography (HPLC). The HPLC chromatographic conditions were as follows: the ratio of acetonitrile to 0.2% phosphoric acid solution was 15 : 85 as the mobile phase. The detection wavelength was 254 nm. The theoretical number of plates should not be less than 3000 based on the peak of ellagic acid. The analysis of ellagic acid content in all RF samples was completed on a C18 column.

(1) Preparation of Control Solution. Take an appropriate amount of ellagic acid standard to accurately weigh, and prepare it into a concentration of $5 \mu g/m$ using 70% methanol as the solvent.

(2) Preparation of the Test Solution. A known amount of RF powder was placed in a 250 ml conical flask with a stopper, followed by the addition of 50 mL of 70% methanol solution. The sample mix was weighted together by thousandth electronic balance with the conical flask and then heated to reflux for 1 hour. Afterward, it was cooled at room temperature ($20°C \pm 5°C$) and filled the reduced weight with 70% methanol. Later, 1 ml of the filtrate was accurately transferred into a 5 ml volumetric flask. Finally, the samples were introduced into an autosampler for HPLC analysis. Determination of ellagic acid content: the injection amounts of the reference solution and the test solution were $10 \mu L$ for everyone.

2.3. Modeling Methods

2.3.1. Spectral Preprocessing Methods. The NIR spectrum is mainly caused by the vibrational absorption of bonds in molecules (C-H, O-H, S-H, N-H), which is the combined and double frequency absorption band of these absorptions [20]. The full spectral band contains a large amount of redundant information, which will increase the operational burden of the PLSR model during building. It will be beneficial to improve the running speed of the model if some irrelevant band information is eliminated in advance. There is litter or no spectral information of the band $9200 \, \text{cm}^{-1} \sim 12000 \, \text{cm}^{-1}$ and no significant feature absorption. Consequently, this spectral region can be removed when establishing the model. In the NIR spectral region, the first and second harmonic generations of OH bonds in moisture molecules exhibit strong absorption bands near 1940 nm ($5154.64 \, \text{cm}^{-1}$) and 1445 nm ($6920 \, \text{cm}^{-1}$), respectively [21]. This can be considered when establishing a quantitative correction analysis model for moisture content. After screening, the modeling spectral regions were $5454.5 \, \text{cm}^{-1} \sim 4598.1 \, \text{cm}^{-1} + 7506.7 \, \text{cm}^{-1} \sim 6094.8 \, \text{cm}^{-1}$, $7506.7 \, \text{cm}^{-1} \sim 4243.2 \, \text{cm}^{-1}$, and $9154.6 \, \text{cm}^{-1} \sim 7499.0 \, \text{cm}^{-1} + 6102.6 \, \text{cm}^{-1} \sim 4243.2 \, \text{cm}^{-1}$ for moisture, total flavonoids, and ellagic acid, respectively.

NIR spectra are generally affected by noise and scattering factors during the period of collection; thus, it is necessary to be preprocessed before modeling. The spectral preprocessing methods included first derivative (1st Der), standard normalized variable (SNV), multiple scatter correction (MSC), elimination constant offset (ECO), and min-max normalization (MMN). 1st Der can remove the constant baseline, which describes the increase or decrease of spectral functions [22]. MSC can effectively eliminate the spectral differences caused by different scattering levels to enhance the correlation between spectra and data [23]. SNV can also correct spectral errors caused by scattering between samples [24]. MMN can remove systematic differences between samples and normalize them to make them comparable to each other [25].

2.3.2. Quantitative Calibration Model. Partial least squares regression (PLSR) is a powerful statistical technique that can construct calibration models for NIR spectral data. It can explore the relationship between independent and dependent variables based on dimensionality reduction techniques [26]. PLSR has been shown its excellent predictive and inference capabilities for NIR spectral modeling [27]. In this work, PLSR was adopted to establish the quantitative calibration model for NIR spectroscopy under the wavelengths ranging from $10000 \, \text{cm}^{-1}$ to $4000 \, \text{cm}^{-1}$. Calibration and test sets according to the $2 : 1$ principle were randomly composed of 67% and 33% of the spectra, respectively. Calibration models for the aforementioned components were established using PLSR. The performance of the models was evaluated by several parameters including correlation coefficient (R), root mean square error (RMSE), and ratio of prediction to deviation (RPD). When R is not less than 0.99, the model is excellent and useable in any application; when R is less than 0.98 and not less than 0.96, the model can be useable in most applications, including quality assurance; when R is less than 0.95 and not less than 0.91, the model can be useable with caution for most applications, including research; when the RPD is less than 0.90 and not less than 0.81, the model can be acceptable for screening and some other "approximate" calibrations; when the RPD is less than 0.80 and not less than 0.71, the model can be acceptable for very rough to rough screening; when the RPD is less than 0.70, the model is poor correlation or even not useable in NIR calibration [28]. The knowledge about the RMSE value is related to the content of components. The closer the RMSE values of calibration and test sets are, the better the predictive ability of the calibration model. When the RPD is less than 1.9, the model is very poor and not recommended; when the RPD is less than 2.4 and not less than 2.0, the model is poor and acceptable for rough screening; when the RPD is less than 2.9 and not less than 2.5, the model is fair and only for screening; when the RPD is greater than or equal to 3.0, the model is good and useable for quality control. Hence, the larger the R and RPD values and the closer the RMSE ratio is to 1, the higher the accuracy of the model and the better its predictive ability [29].

2.3.3. Data Analysis. Each chemical value was detected in triplicate, and the average value was taken. The preprocessing of spectra and the establishment of models were achieved through OPUS 7.0 software. The drawing of graphics was implemented in Origin 2021 software.

3. Results

3.1. NIR Spectra. Under the detailed spectral acquisition indicates of "2.2.1 Sample preparation," the NIR spectra of 97 batches of RF represented by different colors are shown in Figure 1.
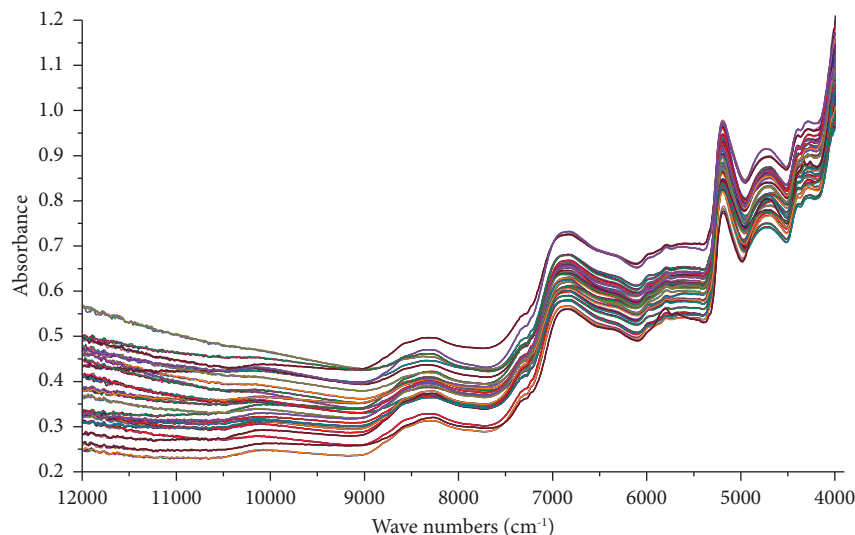
FIGURE 1: The ninety-seven NIR spectra of RF are represented by different colors.

3.2. Contents of Moisture, Total Flavonoids, and Ellagic Acid.

According to the stipulations of the Chinese Pharmacopoeia (2020 edition) for RF, the moisture content shall not exceed 12.0% (General Rule 0832, Second Method) and the ellagic acid ($C_{14}H_6O_8$) content shall not be less than 0.20%. There is no limit on the total flavonoid content. The results of moisture content in RF range from 59 mg/g to 127.6 mg/g (5.90%~12.76%). The results of moisture, total flavonoids, and ellagic acid contents, respectively, determined according to the "2.2.2 Moisture Content," "2.2.3 Total Flavonoids Content," and "2.2.4 ellagic Acid Content" are displayed in Figure 2. Thus, it can be seen that the moisture contents of all experimental RF samples met the requirement (not exceeding 12%), but one was 127.6 mg/g (12.76%) exceeding 12.0% from the production area of Pan'an. The content range of total flavonoids is 20.74 mg/g~63.66 mg/g. The ellagic acid contents range from 6.44 mg/g to 120.08 mg/g (0.64%~12.0%) which were all higher than 0.20%. The ellagic acid content in ninety-seven RF samples was all above 10 mg/g, except for the one that was less than 10% originating from Pan'an. There were two samples higher than 100 mg/g both from Deyang and Lin'an.

3.3. Quantitative Correction Analysis Model

3.3.1. Calibration and Testing Sets.

Ninety-seven experimental samples were randomly divided into calibration and testing sets in a 2:1 ratio. The information on calibration and testing sets for moisture, ellagic acid, and total flavonoids is shown in Table 1.

3.3.2. Building Models.

Near-infrared spectra were modeled by partial least squares regression (PLSR) to predict moisture, total flavonoids, and ellagic acid content. Seven pretreatments were compared during modeling, as follows: nonspectral pretreatment, ECO, MSC, SNV, MMN, 1st
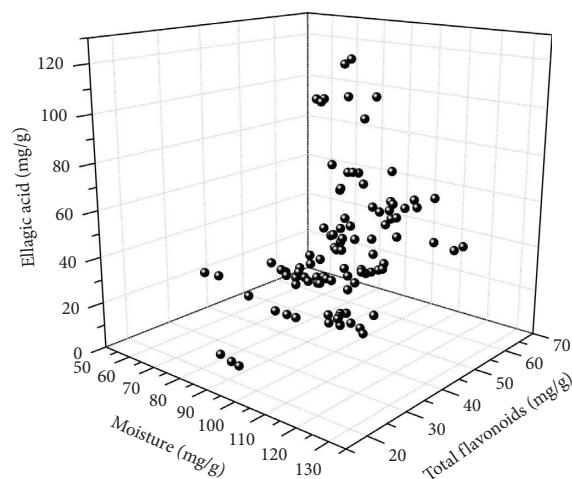

FIGURE 2: Contents of moisture, total flavonoids, and ellagic acid in ninety-seven RF samples.

Der + SNV, and 1st Der + MSC represented by "a," "b," "c," "d," "e," "f," and "g," respectively. PLSR models were optimized for the number of latent variables (LVs) and spectral regions. The results of moisture, total flavonoids, and ellagic acid PLSR models under different pretreatments are shown in Table 2.

The R and RPD values of the moisture PLSR model preprocessed with different pretreatments are all greater than 0.9 and 2.5, respectively. In comparison, there are significant differences between distinct models. The models pretreated with "c," "f," "d," and "g" methods, respectively, have R values of 0.9778 0.9788, 0.9748, and 0.9756, RPD values of 4.8, 4.9, 4.5, and 4.6 as well as RMSE values of 0.306%, 0.300%, 0.321%, and 0.318%, which are better than the R, RPD, and RMSE values of models pretreated with "a," "b," and "e" methods. As a consequence, the models pretreated with "c," "f," "d," and "g" methods have better prediction ability and applicability.

TABLE 1: The information on calibration and testing sets.

Components	Sets	Number	Minimum	Maximum	Average	SD
Moisture (%)	Calibration set	65	6.38	12.76	9.79	1.30
	Testing set	32	6.90	12.18	9.90	1.27
	Total	97	6.38	12.76	9.83	1.29
Total flavonoids (mg/g)	Calibration set	65	20.95	65.41	42.75	8.00
	Testing set	32	21.43	63.66	42.53	8.40
	Total	97	20.95	65.41	42.68	8.09
Ellagic acid (mg/g)	Calibration set	65	6.44	120.08	44.39	26.25
	Testing set	32	7.30	118.72	43.66	27.04
	Total	97	6.44	120.08	44.14	26.38

TABLE 2: Moisture, total flavonoids, and ellagic acid PLSR models with different pretreatments.

Components	Pretreatments	R	RPD	$RMSE_C$	LVs
Moisture (%)	a	0.9571	3.5	0.424	10
	b	0.9503	3.2	0.451	9
	c	0.9778	4.8	0.306	10
	d	0.9788	4.9	0.300	10
	e	0.9202	2.6	0.553	6
	f	0.9748	4.5	0.321	8
	g	0.9756	4.6	0.318	9
Total flavonoids (mg/g)	a	0.9458	3.1	3.09	10
	b	0.9468	3.1	3.07	10
	c	0.9502	3.2	2.97	10
	d	0.9501	3.2	2.97	10
	e	0.8622	2.0	4.55	3
	f	0.8569	1.9	4.60	2
	g	0.8948	2.2	4.01	3
Ellagic acid (mg/g)	a	0.9671	3.9	7.28	10
	b	0.9716	4.2	6.77	10
	c	0.9748	4.5	6.32	9
	d	0.9608	3.6	7.72	7
	e	0.9742	4.4	6.4	9
	f	0.9680	4.0	7.05	8
	g	0.9602	3.6	7.71	6

Note. Description of preprocessing methods: "*a*" represents spectral no pre-treatment, "*b*" represents ECO, "*c*" represents MSC, "*d*" represents SNV, "*e*" represents MMN, "*f*" represents 1st Der + SNV, and "*g*" represents 1st Der + MSC.

As to the total flavonoids PLSR model under different spectral pretreatments, R values have a minimum value of 0.8569 and a maximum value of 0.9502; RPD values range from 1.9 below 2.0 to 3.2 higher; RMSE values range from 2.97 mg/g to 4.60 mg/g. Accordingly, the performance of models with distinct preconditioning methods had significant differences. The R values of models with pretreatments of "*a*," "*b*," "*c*," and "*d*" were 0.9458, 0.9468, 0.9502, and 0.9501, respectively, all greater than 0.9. Furthermore, the RPD values were all greater than 3.2, indicating that those models had good robustness and prediction ability. However, the R values of the models with preprocessing methods "*e*," "*f*," and "*g*" are lower than 0.9. Additionally, the PLSR models of total flavonoids under different spectral pretreatments are generally not as good as the PLSR models of moisture.

The ellagic acid PLSR models under different spectral pretreatments are generally better than the moisture and total flavonoids PLSR models, with R values all greater than 0.96 and RPD values greater than 3.5. After careful comparison, the models preprocessed with "*b*," "*c*," "*e*," and "*f*" methods are more outstanding, whose R values are 0.9716, 0.9748, 0.9742, and 0.9680, RPD values are 4.2, 4.5, 4.4, and 4.0 as well as RMSE values are 6.77 mg/g, 6.32 mg/g, 6.4 mg/g, and 7.05 mg/g, respectively. In contrast, the performance parameters with "*a*," "*d*," and "*g*" methods were slightly inferior, for instance, RPD values below 4.0.

3.3.3. Validating Models. According to "3.3.2 Building Models," the top three moisture, total flavonoids, and ellagic acid PLSR models were obtained for validation. As shown in Table 3, there superior moisture PLSR models were preprocessed with "*c*," "*d*," "*f*," and "*g*." After testing, it should be noted that the model under the "d" method preprocessing is the best with R values above 0.96 and RPD values greater than 3.5. The ratio of RMSE values between the calibration set and the test set is 0.98 closer to 1. In comparison, the other models are slightly inferior, either R less than 0.96 or RPD lower than 3.5. The top four PLSR models of total flavonoids with "*a*," "*b*," "*c*," and "*d*" pretreatments were validated by the testing set. The results are not very satisfactory, for the R values are greater than 0.85 but less than 0.9, as well as the RPD values range from 1.9 to 2.3. Anyway, the total flavonoids PLSR model with "*b*" pretreatment was slightly better, with a smaller RMSE value of 2.90 mg/g, and higher R and RPD values of 0.8981 and 2.3, respectively. Four PLSR models of ellagic acid preprocessed with "*b*," "*c*," "*e*," and "*f*" pretreatments were employed to test and verify. The R values were all above 0.96, RPD values were 3.4, 4.2, 4.0, and 3.9, and RMSE values were 7.34 mg/g, 6.32 mg/g, 6.74 mg/g, and 6.84 mg/g, respectively. Overall, the best performance of ellagic acid model was preprocessed by the "*c*" method.

3.3.4. PLSR Models. The moisture model pretreated with SNV method is shown in Figure 3(a), which revealed the relationship between observed and predictive values. The R-value was 0.9788, indicating a good correlation between predicted and chemical values. The validated result of the

TABLE 3: Validation of the moisture, total flavonoids, and ellagic acid PLSR models.

Components	Pretreatments	R	RPD	RMSE_P	(RMSE_C/RMSE_P)
Moisture (%)	c	0.9490	3.2	0.365	0.84
	d	0.9644	3.8	0.307	0.98
	f	0.9578	3.4	0.333	0.96
	g	0.9482	3.2	0.368	0.86
Total flavonoids (mg/g)	a	0.8505	1.9	3.09	0.88
	b	0.8981	2.3	2.90	0.94
	c	0.8853	2.1	3.05	0.97
	d	0.8826	2.1	3.06	0.97
Ellagic acid (mg/g)	b	0.9612	3.4	7.34	0.92
	c	0.9714	4.2	6.32	1.00
	e	0.9674	4.0	6.74	0.95
	f	0.9692	3.9	6.84	0.97

Note. Description of preprocessing methods: "*a*" represents spectral free pretreatment, "*b*" represents ECO, "*c*" represents MSC, "*e*" represents MMN, "*f*" represents 1st Der + SNV, and "*g*" represents 1st Der + MSC.
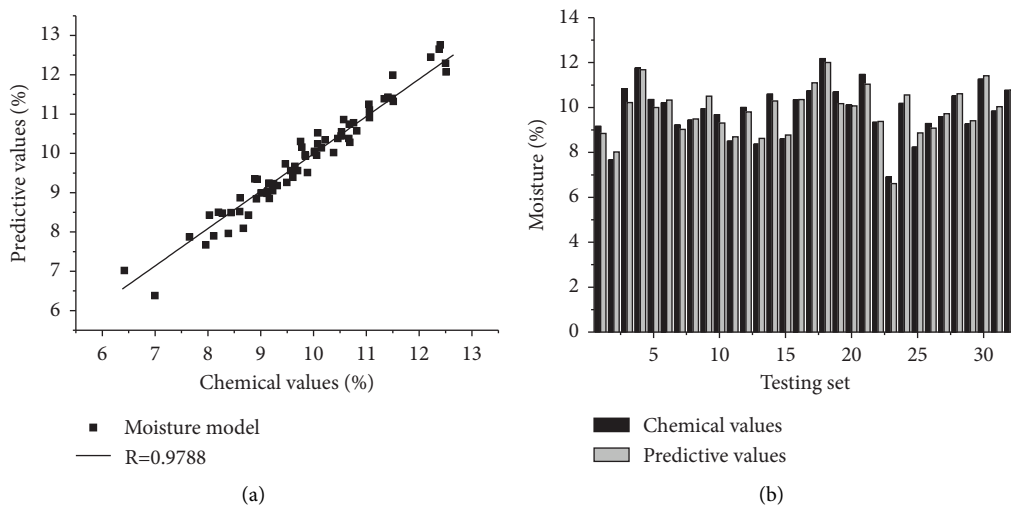

(a)

(b)

FIGURE 3: PLSR model for moisture (a) and its testing results (b).
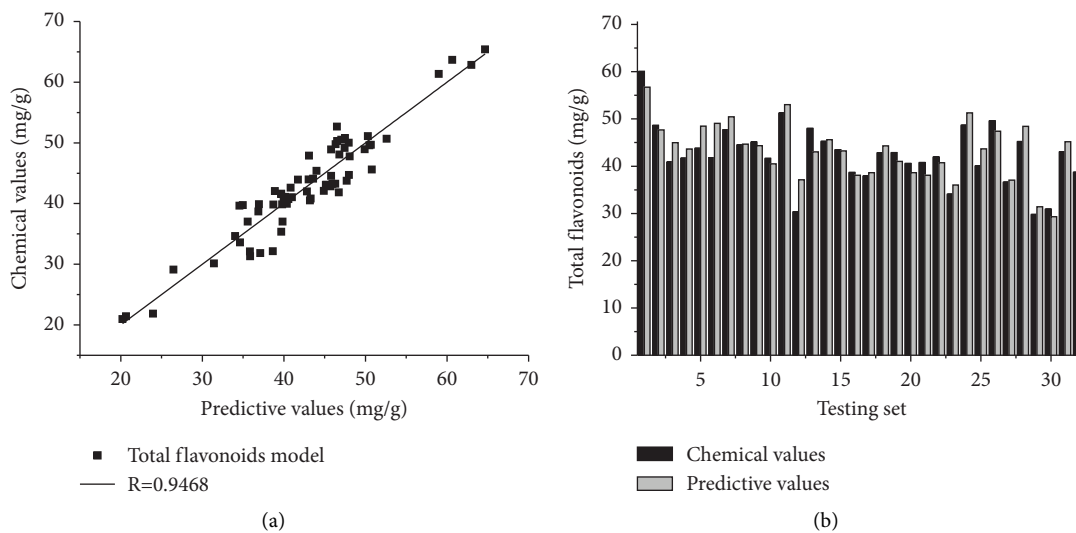

(a)

(b)

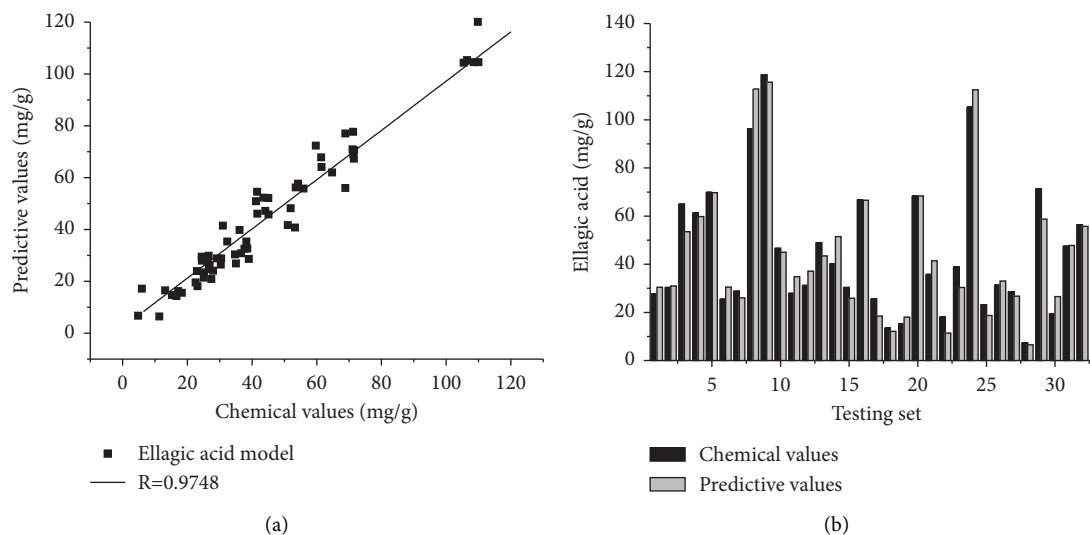FIGURE 4: PLSR model for total flavonoids (a) and its testing results (b).

Figure 5: PLSR model for ellagic acid (a) and its testing results (b).

model possessed an RPD value of 3.8 larger than 3.5. The predicted value was relatively close to the chemical value, indicating that the model had good prediction accuracy, as shown in Figure 3(b).

The calibration model for total flavonoids is shown in Figure 4(a), with ECO as the best preprocessing method, an R-value of 0.9468 indicating a good correlation between predicted and the chemical values. The verified results of the model by the test set are shown in Figure 4(b), with an RPD value of 2.3 and an RMSE ratio of 0.94 between the calibration and the testing sets, indicating that the established quantitative model can be useable for rough screening.

The calibration model for the quantitative analysis of ellagic acid is shown in Figure 5(a). The best pretreatment method was MSC, and the R value was 0.9748 indicating that the model had good correlation. The verified results by the testing set are shown in Figure 5(b), with an RPD value of 4.2 and an RMSE ratio of 1 between the calibration and testing sets, indicating that the established model has quite good prediction ability.

4. Conclusion

This work measured the moisture, total flavonoids, and ellagic acid content of ninety-seven RF samples from twenty distinct geographic regions. Among them, samples from Pan'an had an abnormal moisture content of 127.6 mg/g (12.76%), which was exceeding 12.0%. The remaining RF samples were all within 12% for moisture content. The total flavonoid content ranged from 20.74 mg/g to 63.66 mg/g. The content range of ellagic acid was 6.44 mg/g~120.08 mg/g (0.64%~12.0%), all higher than 0.2%. Ellagic acid content varying the largest among distinct geographic regions, followed by total flavonoids and finally moisture content.

Additionally, the ellagic acid content and moisture levels can essentially meet the standards of the Chinese Pharmacopoeia (2020 edition).

Those quantitative analysis calibration models for moisture, total flavonoids, and ellagic acid content in RF were established through NIR spectroscopy combined with PLSR chemometric. By comparing seven preprocessing methods, the best spectral preprocessing method was selected. It was proven that suitable preprocessing methods could improve the predictive ability through the testing sets. The finally selected PLSR models all had good predictive performance, and the predicted results were close to the measured values. The results obtained in this work contributed to increasing the knowledge about the ability of PLSR models to quickly evaluate the RF quality. The predicted values of moisture, total flavonoids, and ellagic acid in RF are relatively reliable.

The inadequacy of this work is that a landmark component named kaempferol-3-O-rutoside was not considered to establish the PLSR model owing to its value generally greater than 0.03% but less than 0.1%. A little kaempferol-3-O-rutoside in RF does not meet the detection limit of a universal near-infrared spectrometer (≥0.1%). Therefore, in the later stage, it is possible to establish and achieve a rapid PLSR model for low-content components.

Data Availability

The xls. data used to support the findings of this study were supplied by Chunyan Wu under license and so cannot be made freely available. Requests for access to these data should be made to Chunyan Wu, wcy13575460036@126.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] China Medical Science and Technology Press, *National Pharmacopoeia Committee Pharmacopoeia of the People's Republic of China: Part One*, China Medical Science and Technology Press, Beijing, China, 2020.

[2] Q. Q. Chen, K. Li, X. Q. Tang, L. Geng, L. Wang, and Y. P. Peng, "Analysis of phenolic components and antioxidant activity in the fruits, stems, and leaves of East China raspberry," *Food Science*, vol. 41, pp. 209–215, 2020.

[3] J. Zhong, Y. Wang, C. Li et al., "Natural variation on free, esterified, glycosylated and insoluble-bound phenolics of *Rubus chingii Hu*: correlation between phenolic constituents and antioxidant activities," *Food Research International*, vol. 162, Article ID 112043, 2022.

[4] Y. Huo, X. Zhao, J. Zhao et al., "Hypoglycemic effects of Fu-Pen-Zi (*Rubus chingii Hu*) fruit extracts in streptozotocin-induced type 1 diabetic mice," *Journal of Functional Foods*, vol. 87, Article ID 104837, 2021.

[5] L. L. Cui, X. M. Yuan, Y. H. Gao, J. F. Jiang, and Y. X. Guan, "Study on the mechanism of raspberry in treating type 2 diabetes based on UPLC-Q-Orbitrap-MS and network pharmacology," *Global Traditional Chinese Medicine*, vol. 15, pp. 1570–1579, 2022.

[6] X. Y. Su, M. Zhao, X. Y. Fu, X. D. Ma, W. Xu, and S. H. Hu, "Immunomodulatory activity of purified polysaccharides from *Rubus chingii Hu* fruits in lymphocytes and its molecular mechanisms," *Journal of Functional Foods*, vol. 87, Article ID 104785, 2021.

[7] H. H. Ke, T. Bao, and W. Chen, "Polysaccharide from *Rubus chingii Hu* affords protection against palmitic acid-induced lipotoxicity in human hepatocytes," *International Journal of Biological Macromolecules*, vol. 133, pp. 1063–1071, 2019.

[8] J. Bauza-Kaszewska, E. Żary-Sikorska, A. Gugolek et al., "Synergistic antimicrobial effect of raspberry (*Rubus idaeus* L., rosaceae) preparations and probiotic bacteria on enteric pathogens," *Polish Journal of Food and Nutrition Sciences*, vol. 71, pp. 51–59, 2021.

[9] Z. H. Liu, Z. Mi, N. Han, J. Zhai, S. Li, and J. Yin, "Diterpenoid glucosides with anti-inflammatory activity from Rubi Fructus," *Fitoterapia*, vol. 164, Article ID 105325, 2023.

[10] J. M. Zou, *Study on Phenolic Active Ingredients against Alzheimer's Disease in Raspberries Based on Molecular Imprinting and Docking techniques*, Jiangxi University of Traditional Chinese Medicine, Nanchang, China, 2021.

[11] J. Tan, D. S. Wang, Y. Lu et al., "Metabolic enzyme inhibitory abilities, in vivo hypoglycemic ability of palmleaf raspberry fruits extracts and identification of hypoglycemic compounds," *Food Science and Human Wellness*, vol. 12, no. 4, pp. 1232–1240, 2023.

[12] X. Yao, F. C. Mao, J. B. Cheng, C. L. Yu, and W. W. Yu, "Establishment of UPLC analysis system for raspberry palm leaf and dynamic analysis of medicinal components in green fruits," 2022, http://kns.cnki.net/kcms/detail/46.1068.S.20220719.1105.004.html.

[13] Y. Jiang, T. Luo, Y. Tang et al., "Isolation of a novel characterized *Issatchenkia terricola* from red raspberry fruits on the degradation of citric acid and enrichment of flavonoid and volatile profiles in fermented red raspberry juice," *Food Science and Human Wellness*, vol. 11, no. 4, pp. 1018–1027, 2022.

[14] Z. M. Zhou, D. M. Yan, Y. K. Wang et al., "Discovery of quality markers in *rubus chingii hu* using UPLC-ESI-QTOF-MS," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 203, Article ID 114200, 2021.

[15] Y. Q. Wu, C. H. Zhang, Z. J. Huang et al., "The color difference of *rubus* fruits is closely related to the composition of flavonoids including anthocyanins," *Lebensmittel-Wissenschaft and Technologie*, vol. 149, Article ID 111825, 2021.

[16] A. J. Yan, W. T. Zhang, Y. B. Zhou, S. L. Yu, Y. M. Zhou, and Y. Hu, "Identification of fu-pen-zi and raspberry based on gas chromatography ion migration spectroscopy," *Journal of Chinese Medicinal Materials*, vol. 45, pp. 574–578, 2022.

[17] A. Tantinantrakun, S. Sukwanit, A. K. Thompson, and S. Teerachaichayut, "Nondestructive evaluation of SW-NIRS and NIR-HSI for predicting the maturity index of intact pineapples," *Postharvest Biology and Technology*, vol. 195, Article ID 112141, 2023.

[18] Z. W. Huang, T. Guo, W. J. Huang et al., "Research progress of near-infrared spectroscopy in quality evaluation of valuable Chinese medicinal materials," *Chinese Traditional and Herbal Drugs*, vol. 53, pp. 6328–6336, 2022.

[19] Y. F. Wu, J. Wu, and W. W. Yu, "Optimization of ultrasonic extraction process of total flavonoids from raspberry by response surface methodology and study on its antioxidant characteristics," *Journal of Zhejiang Forestry Science and Technology*, vol. 34, pp. 23–27, 2014.

[20] C. Pasquini, "Near infrared spectroscopy: fundamentals, practical aspects and analytical applications," *Journal of the Brazilian Chemical Society*, vol. 14, no. 2, pp. 198–219, 2003.

[21] M. Streefland, P. F. Van Herpen, B. Van de Waterbeemd et al., "A practical approach for exploration and modeling of the design space of a bacterial vaccine cultivation process," *Biotechnology and Bioengineering*, vol. 104, no. 3, pp. 492–504, 2009.

[22] G. Jiang, K. Zhou, J. Wang et al., "Estimation of rock copper content based on Fractional-order derivative and visible Near-infrared-Shortwave infrared spectroscopy," *Ore Geology Reviews*, vol. 150, Article ID 105092, 2022.

[23] T. P. Wrobel, D. Liberda, P. Koziol, C. Paluszkiewicz, and W. M. Kwiatek, "Comparison of the new mie extinction extended multiplicative scattering correction and resonant mie extended multiplicative scattering correction in transmission infrared tissue image scattering correction," *Infrared Physics & Technology*, vol. 107, Article ID 103291, 2020.

[24] X. Y. Duan, *Research on Quality Analysis and Control of Forsythia Suspensa*, Henan University of Traditional Chinese Medicine, Zhengzhou, China, 2014.

[25] W. Li and Z. Liu, "A method of SVM with normalization in intrusion detection," *Procedia Environmental Sciences*, vol. 11, pp. 256–262, 2011.

[26] H. T. Shi and P. Q. Yu, "Comparison of grating-based near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/

MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat," *Food Control*, vol. 82, pp. 57–65, 2017.

[27] H. Yan, Z. J. Pu, Y. Wang et al., "Rapid qualitative identification and quantitative analysis of Flos Mume based on Fourier transform near infrared spectroscopy," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 249, Article ID 119344, 2021.

[28] P. Williams, J. Antoniszyn, and M. Manley, *Near Infrared Technology: Getting the Best Out of light*, African Sun Media, Stellenbosch, South Africa, 1st edition, 2019.

[29] X. W. Chang, D. D. Wei, S. L. Su et al., "An integrated strategy for rapid discovery and prediction of nucleobases, nucleosides and amino acids as quality markers in different flowering stages of Flos *Chrysanthemi* using UPLC-MS/MS and FT-NIR coupled with multivariate statistical analysis," *Microchemical Journal*, vol. 153, Article ID 104500, 2020.