

## Research Article

# Quantitative Determination of Sucrose Adulterated in Red Ginseng by Terahertz Time-Domain Spectroscopy (THz-TDS) with Monte Carlo Uninformative Variable Elimination (MCUVE) and Support Vector Regression (SVR)

Shubao Pan,<sup>1,2</sup> Huo Zhang ,<sup>1</sup> Zhi Li ,<sup>2</sup> Tao Chen,<sup>1</sup> and Xianhua Yin<sup>1</sup>

<sup>1</sup>Guangxi Key Laboratory of Automatic Detecting Technology and Instruments,  
School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

<sup>2</sup>Guilin University of Aerospace Technology, Guilin, Guangxi 541004, China

Correspondence should be addressed to Huo Zhang; zhanghuo@outlook.com and Zhi Li; cclizhi@guat.edu.cn

Received 20 October 2021; Revised 8 January 2022; Accepted 28 January 2022; Published 27 February 2022

Academic Editor: Jose M. Pedrosa

Copyright © 2022 Shubao Pan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper introduces a method to detect the content of sucrose, an adulterant of red ginseng, based on terahertz spectroscopy. Experiments were carried out on red ginseng with 6 levels of adulterated concentrations using terahertz time-domain spectroscopy (THz-TDS). We separately extracted the information of the terahertz spectral curve by principal component analysis (PCA) and Monte Carlo uninformative variable elimination (MCUVE) and then separately performed quantitative analysis by partial least squares regression (PLSR) and support vector regression (SVR). Because the nonlinear line factor in the terahertz spectral curve of red ginseng samples is considered, the MCUVE-SVR has high correlation coefficient ( $>0.99$ ) and ratio prediction to deviation ( $>7.4$ ), low root means square error of deviation ( $<1.2\%$ ), and Bias ( $<0.05\%$ ). The results prove that MCUVE-SVR can be regarded as an ideal quantitative analysis method in the detection of sucrose incorporation in red ginseng by terahertz spectroscopy.

## 1. Introduction

In Asia, red ginseng is a kind of famous herb that is popular among people because of its positive effects on cognitive ability [1], anti-ageing [2], anti-oxidation [3], anti-inflammatory [4], anti-obesity [5], and improving immunity [6]. However, high-quality red ginseng is not easy to process. First of all, qualified fresh ginseng needs to be picked out, washed, and then steamed at a specified time and dried into red ginseng finally [7]. In order to obtain more benefits, some unlawful businessmen add cheap sucrose during the processing of red ginseng, which can increase the weight and improve the color [8]. Therefore, an effective method to detect whether red ginseng is adulterated is necessary to ensure its quality.

So far, the most commonly used methods for the detection of sucrose are chemical analysis [9] and high-performance

liquid chromatography (HPLC) [10–12]. At the same time, there are some other methods suitable for quantitative analysis, such as liquid chromatography–mass spectrometry (LC-MS) [13], nuclear magnetic resonance (NMR) [14], and capillary electrophoresis (CE) [15]. All these methods can obtain accurate and objective results but often require additional chemical consumption and complex operations. The result is not only the rising cost and time consumption but also the detected red ginseng can no longer be used. In East Asian countries, precious red ginseng is in great demand, so a rapid and nondestructive detection technology is valuable.

From many research reports, terahertz time-domain spectral systems (THz-TDS) have been applied to research on drugs [16], food safety [17], materials [18], and biomedicine [19, 20], and the results show that terahertz time-domain spectroscopy technology is a fast, nondestructive, and nonpolluting detection technology.

For the analysis of component content of terahertz spectroscopy, support vector regression (SVR) and partial least-squares regression (PLSR) are the two most commonly used algorithms, and they are used in applications such as evaluation of peroxide value in peanut oils [21], detection of maltose in wheat grains [22], and detection of octogen content [23]. The Monte Carlo uninformative variable elimination (MCUVE) was originally an effective information extraction algorithm for the application of PLSR [24, 25]. To obtain a potential ideal way for the detection of adulteration of red ginseng, we further consider the non-linearity and the removal of invalid data on the basis of studying SVR and combined SVR with MCUVE.

## 2. Materials and Methods

**2.1. Sample Preparation.** Red ginseng in dried form was purchased from Tongrentang Co., Ltd. (Beijing, China), and the sucrose was purchased from Aladdin Biochemical Technology Co., Ltd. (Shanghai, China). The purity of sucrose is above 99.9%.

Since herbs are not a single component and the spatial distribution of components is uneven, the sample needs to be ground into powder to mix evenly. Because there is no chemical reaction in the testing process, the tested samples will not lose their medicinal value. Therefore, the detection method based on terahertz spectroscopy is considered nondestructive.

The red ginseng was crushed into powder with a grinder (DFX-X200, Wenzhou Dingli Medical Instrument Co., China), while the sucrose was ground into powder with a pestle and mortar. All powders were sieved with 200 meshes and then placed in a drying box, where they were dried at a temperature of 50°C for two hours to remove water. Each sample is made independently. The powders of red ginseng were added with sucrose powders and then mixed evenly at a variety of concentrations of 5%, 10%, 15%, 20%, 25%, and 30%. The appearance change of red ginseng is obvious when the sucrose concentration is more than 30% and can be easily identified by observation. After that, the powders were made into circular tablets with a diameter of 13 mm and a thickness of 1.2 mm under a pressure of 12 MPa by using a hydraulic press (PC-15, Tianjin Jingtuo Instrument Technology Corp., China). The weight of each sample is about 220 mg. There are 36 samples made for each concentration. For each sucrose content, samples were randomly selected in the form of sampling without replacement, until 24 samples were selected. These 24 samples were put into the training set, and the remaining 12 samples were used as the testing set. Repeat this operation until each sucrose content has been sampled. Then, the training set has 144 samples and the testing set has 72 samples.

**2.2. Instrumentation.** The terahertz time-domain spectral system (THz-TDS) used in this study is composed of a terahertz time-domain spectrometer (Z3, Zomega Terahertz Corp., USA) and a femtosecond laser (FemtoFiber pro NIR, TOPTICA Photonics Inc., Germany). The structure is shown

in Figure 1. When the laser is working, it generates laser pulses with a repetition frequency of 82 Hz, a pulse width of about 100 fs, a wavelength center of 780 nm, and an average power of nearly 100 mW. The laser beam is divided into a pump beam and a probe beam through a cube beam splitter (CBS). The terahertz pulse generated by the pump beam irradiating the photoconductive antenna penetrates the sample under test. After passing through the delayed optical path, the probe light and the terahertz pulse pass through the ZnTe crystal, a quarter-wave plate (QWP), and a polarization-beam-splitter (PBS) and then irradiate on the detector. The beam path of the terahertz time-domain spectrometer is placed in a closed box and injected with dry air. To ensure the accuracy of the experiment, the terahertz spectrum of the samples was collected only when the humidity in the box was less than 1% and the indoor temperature was 25°C.

**2.3. Data Acquisition.** The dry air is tested to obtain the reference signal and the samples are tested to obtain the sample signals. The signals are calculated by fast Fourier transform (FFT), and the reference spectrum  $E_{ref}(\omega)$  and the sample signal spectrum  $E_s(\omega)$  are obtained, respectively.  $\omega$  is the sample frequency. Finally, the terahertz absorption spectrum can be calculated, which can reflect the absorption level of the sample. The calculation formula is shown in the following formula:

$$\text{absorbance}(\omega) = -\log_{10} \left| \frac{E_s(\omega)}{E_{ref}(\omega)} \right|^2. \quad (1)$$

## 3. Modeling Approach

**3.1. Principal Component Analysis (PCA).** The terahertz absorption spectrum of the samples contains a wealth of information, which also contains redundant useless information. Therefore, an algorithm that retains effective information as much as possible and can simplify the data is needed. PCA is the most widely used dimensionality reduction algorithm. By transforming the coordinate system, the algorithm projects the high-dimensional original data into the low-dimensional data space and ensures the maximum variance of the reserved data. Generally speaking, the first few important principal components already contain most of the information. PCA simplifies the original data, which helps reduce the computational complexity of the model, and the accuracy of the model is almost unaffected [26].

Ni et al. studied the selection basis of principal components (PCs) [27]. Firstly, the standard deviation (SD) is calculated as follows:

$$\text{S.D.} = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^p \epsilon_{ij}^2}{(n-1)(p-m)}}, \quad (2)$$

where  $\epsilon_{ij}$  is the squared residual of an object  $i$  on principal component  $j$ ,  $n$  is the number of objects,  $p$  is the number of

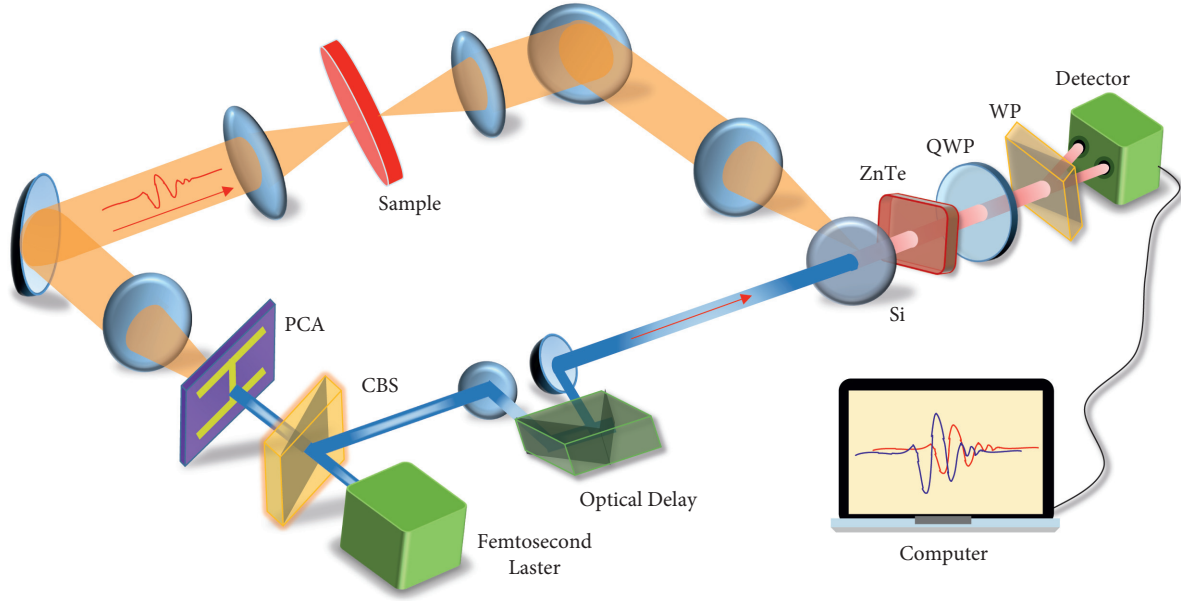


FIGURE 1: Schematic diagram of THz-TDS spectrometer.

data in the detection band, and  $m$  is the number of principal components.

Then, the noise level (NL) of the instrument is calculated. It can be calculated using the following formula by measuring the sample  $q$  times:

$$NL = \sqrt{\frac{\sum_{k=1}^q \sum_{l=1}^p (x_{kl} - \bar{x}_l)^2}{p(q-1)}}, \quad (3)$$

where  $x_{kl}$  is the absorption spectrum,  $k$  is the number of scans,  $l$  is the wavelength, and  $\bar{x}_l$  is the average absorption spectrum at the wavelength  $l$ , which is calculated as follows:

$$\bar{x}_l = \frac{1}{q} \sum_{k=1}^q x_{kl}. \quad (4)$$

In order to determine the upper limit on the number of principal components, the cut off to determine maximum number of PCs is defined as

$$r = \frac{\text{S.D.}}{\text{NL}}. \quad (5)$$

When the  $r$  value is 1.5, the corresponding number of PCs is considered ideal.

**3.2. PLSR.** Owing to its good analytical performance and robustness, partial least squares regression PLSR has become very popular in multiple regression analysis applications [28, 29]. This multiple regression analysis methods constructs the relationship between the predicted concentration and the spectral matrix, as shown below

$$y = Xb + e, \quad (6)$$

where  $y$  is the concentration matrix,  $X$  is the spectrum matrix,  $b$  is the regression coefficient vector, and  $e$  is the error vector [30].

PLSR can perform regression analysis on multiple variables, with low computational complexity and fast data processing speed. The algorithm has more advantages when the data features are relatively linear. For the nonlinear situation, the partition and variable selection should be optimized; otherwise, the prediction accuracy will be affected.

**3.3. Monte Carlo Uninformative Variable Elimination (MCUVE).** The UVE (uninformative variable elimination) algorithm is a typical variable screening method, which can eliminate uninformative variables and reduce errors caused by these variables. Based on the PLS model, the UVE algorithm adds a certain number of random variable matrices and obtains a regression coefficient matrix  $B$  by using leave-one-out cross-validation. The stability  $C_j$  of the coefficient  $B$  is defined as

$$C_j = \frac{\text{mean}(B_j)}{\text{std}(B_j)}, \quad \text{for } j = 1, \dots, P, \quad (7)$$

where  $\text{mean}(B_j)$  is the average value of the regression coefficient  $B_j$  and  $\text{std}(B_j)$  is the variance of the regression coefficient  $B_j$ . Let  $C_{\text{artif}}$  be the maximum value obtained by the artificial variable. When  $\text{ABS}(C_j)$  is less than  $\text{ABS}(C_{\text{artif}})$ , the variable is eliminated [31].

Monte Carlo cross-validation can be used in the original UVE algorithm to replace leave-one-out cross-validation. This requires multiple training and verification of the model, and each time the data is randomly divided into a training set and a verification set. Finally, all the results are averaged as

the verification error of the model. Making such changes reduces data dependence and improves the stability of the model [32].

**3.4. SVR.** SVR is also a very common regression analysis model. Different from the SVM classification method, SVR constructs the optimal hyperplane and then adds the variable  $\varepsilon$  to form the upper and lower planes parallel to the hyperplane. The two planes sandwich the sample points in the middle while minimizing the distance between the two planes. With the help of kernel functions, the algorithm maps low-dimensional spatial data to high-dimensional spatial data, which can solve the nonlinear problem between independent variables and dependent variables. Due to its powerful linear and nonlinear regression analysis capabilities, SVR regression methods have been successfully used in some applications [33, 34].

In this paper, radial basis function (RBF) was used as the kernel function of the SVR model and is expressed as follows:

$$\kappa(x_i, y_i) = \exp\left(-\frac{\|x_i - y_i\|^2}{\gamma^2}\right), \quad (8)$$

where  $x_i$  is the point in space and  $y_i$  is the center of the kernel function. The genetic algorithm, an excellent optimization algorithm commonly used in SVR, was used to search the optimal penalty factor  $C$  of the SVR model and the optimal parameter  $\gamma$  of RBF.

**3.5. Model Validation.** The performance of model prediction is generally determined by the correlation coefficient ( $R$ ), the root mean square deviation (RMSD), the Bias, and the ratio of prediction to deviation (RPD). Their definitions are as follows:

$$\begin{aligned} R &= \frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}}, \\ \text{RMSD} &= \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - 1}}, \\ \text{Bias} &= \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i), \\ \text{RPD} &= \frac{1}{\sqrt{1 - R^2}}. \end{aligned} \quad (9)$$

Here,  $N$  is the number of samples,  $y_i$  is the reference concentration of the  $i$ -th sample,  $\bar{y}$  is the average of the reference concentration,  $\hat{y}_i$  is the predicted concentration of the  $i$ -th sample, and  $\bar{\hat{y}}$  is the average of the predicted concentration.

We used 10-fold cross-validation to obtain the validation set from the training set, whose correlation coefficient and the root mean square error indicate the performance of the model and are expressed as  $R_v$  and

RMSDV, respectively. Correspondingly, the correlation coefficient and root mean square error of the testing set indicate the generalization ability of the model and are expressed as  $R_p$  and RMSDP, respectively. The closer the  $R$  value less than 1 and the smaller the root mean square error, the better the model.

In the final evaluation of the model, in addition to  $R$  and RMSE, Bias and RPD will also be combined. Bias indicates the overall deviation between the predicted value and the actual value, so the smaller the number, the better. RPD indicates the predictive ability of the model. When the RPD is greater than 3.5, the model is considered ideal. If the value of RPD is greater, the model is considered better.

## 4. Results and Discussion

**4.1. Spectral Analysis.** The time-domain spectra of red ginseng and sucrose samples are obtained by THz-TDS and are shown in Figure 2. The corresponding absorbance spectra in the band of 0.3–1.6 THz were calculated from the time-domain spectra and are presented in Figure 3. It could be seen that four chief absorption peaks of red ginseng could be found at 1.03, 1.17, 1.13 and 1.45 THz, while the chief absorption peak of sucrose was at 1.47 THz. According to Lambert–Beer’s law, absorbance is proportional to the concentration and thickness of the sample. In the case of our quantitative analysis, the thickness is fixed, so what we are interested in are the changes in the absorbance curves. As shown in Figure 4, when red ginseng and sucrose are mixed in different ratios, the main absorption peaks can also be observed and change according to the content. However, a large number of curves and the mutual interference between curves make it impossible to find the change law through observation. In order to obtain the content ratios, pattern recognition methods are usually used to establish an analysis model.

**4.2. Feature Selection.** In order to deal with redundant information and computation, it is necessary to reduce the dimension of spectral data. Before dimensionality reduction, in order to eliminate the influence of dimension and value range differences between indexes, the Z-score was used for the standardization of spectral data. Then, the spectral curves were transformed into sets of linearly uncorrelated variables by PCA, and the results are shown in Figure 5. According to Ni et al. [27], in order to achieve a balance between reducing interference and information loss, the value of  $r$  in formula (5) is set to 1.5, and the calculated number of PCs is 30.

Another dimension reduction method used in this paper is MCVUE. This method tries to remove the invalid information carried in the spectral data, reduces the data scale within a reasonable limit, and does not affect the amount of effective information carried by the data. Figure 6 shows the results of the stability of subsets of spectral bands selected by MCVUE. In order to compare with PCA, we selected the first 30-dimensional data with the highest stability to establish the model.

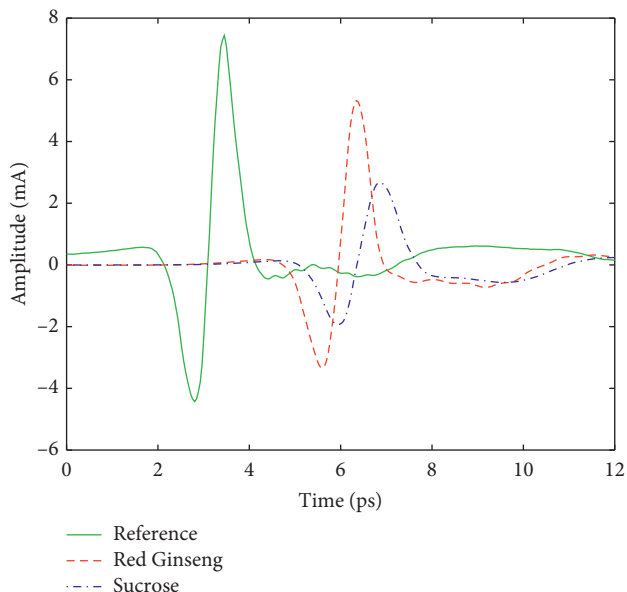


FIGURE 2: The time-domain spectra of red ginseng and sucrose.

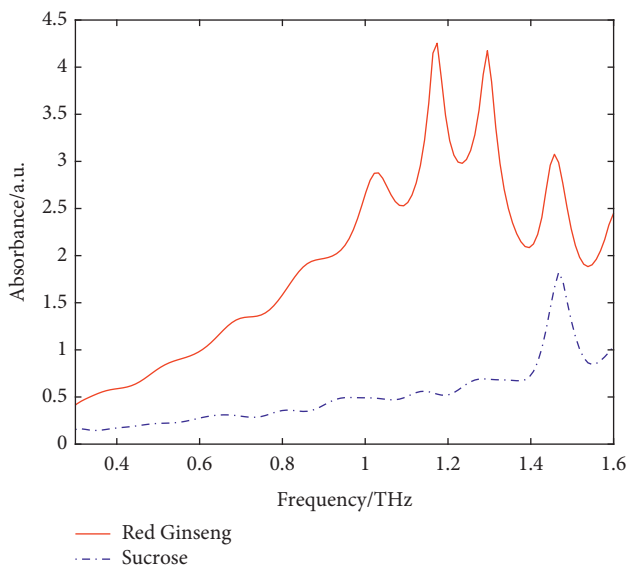


FIGURE 3: The average absorbance spectra of red ginseng and sucrose.

4.3. *Quantitative Determination.* The quantitative analysis results of different models are shown in Table 1. The relationship between the observed sucrose content and the results predicted by each regression model is shown in Figure 7. The accuracy of the models is judged by comparing RMSDV,  $R_v$ , RMSDP, and  $R_p$ .

According to Beer-Lambert's Law, it can be considered that the spectrum of the mixture is the linear superposition of the spectra of each component. However, Beer-Lambert's Law is a limited law, and there are some factors, such as interaction between components of the mixture and the disturbance of stray light, that invalidate it in the actual spectrum acquisition process. As a result, the nonlinear factors of the herbs' terahertz spectra are often not

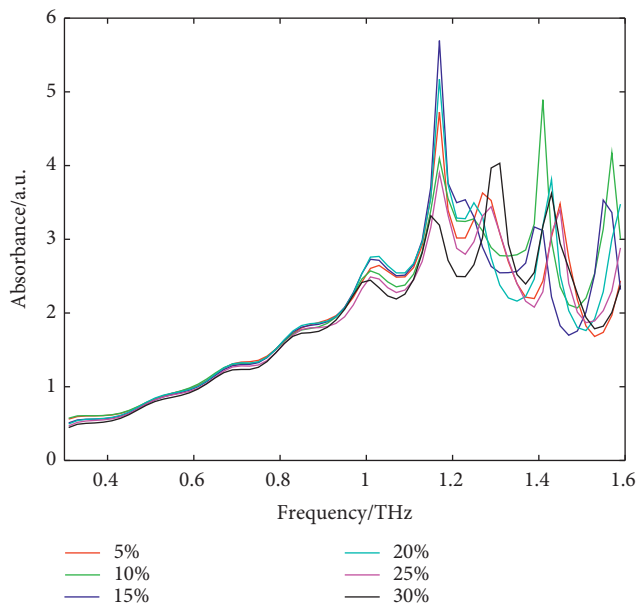


FIGURE 4: The average absorption spectra of red ginseng and sucrose mixed in different ratios.

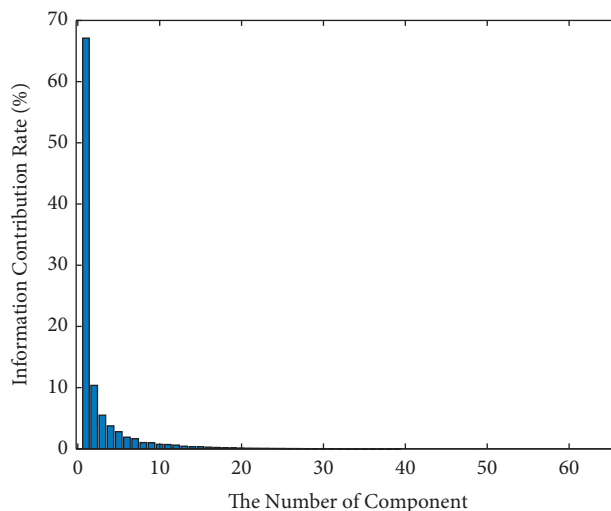


FIGURE 5: The information contribution rates of the data processed by PCA.

negligible. Theoretically, SVR, a method with strong nonlinear analysis ability, can predict the sucrose content in red ginseng more accurately than PLSR, a linear regression method. The results show that SVR obtains better prediction results than PLSR, which is manifested in larger  $R_p$  and smaller RMSDP. Meanwhile, PLSR has a larger  $R_v$  and a smaller RMSDV than SVR. Better training set analysis results and worse testing set analysis results mean that PLSR has less fit.

In the case of combination with PCA,  $R_v$  and  $R_p$  of PLSR remain at 0.988 and 0.985, respectively. RMSDV increased by 0.015%, and RMSDP decreased slightly. This shows that compared with a single PLSR, the overfitting of PCA-PLSR is not improved and the prediction ability is also not enhanced. Combined with the same feature

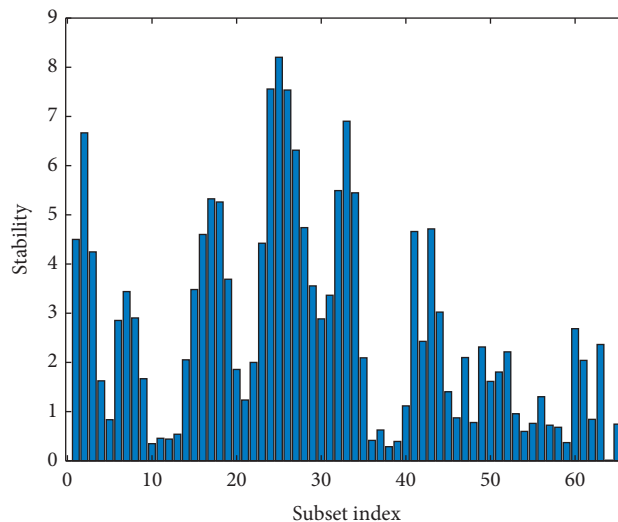


FIGURE 6: The stability of the data processed by MCUVE.

TABLE 1: Performance of the models.

Model	Rv	RMSDV (%)	Rp	RMSDP (%)
PLSR	0.988	1.323	0.985	1.500
SVR	0.987	1.402	0.987	1.494
PCA-PLSR	0.988	1.338	0.985	1.499
PCA-SVR	0.989	1.272	0.987	1.412
MCUVE-PLSR	0.987	1.388	0.986	1.422
MCUVE-SVR	0.993	0.999	0.991	1.172

extraction method, Rv of SVR increased from 0.987 to 0.989 and Rp remains at 0.987. RMSDV and RMSDP decreased in varying degrees. It can be seen that the noise reduction effect on the terahertz spectrum of red ginseng samples is not obvious, and the improvement brought by PCA-SVR is very limited.

While PCA reduces noise, effective information will also be reduced. When the advantages brought by noise reduction cannot fill the loss brought by effective information reduction, it will have an adverse impact on the prediction ability of the model. After replacing PCA with MCUVE, both PLSR and SVR achieved better performance. Especially, with the combination of SVR and MCUVE, RMSDV and RMSDP are also reduced to the lowest error values of these models Rv increased from 0.989 to 0.993, and Rp increased from 0.987 to 0.990; RMSDV decreased from 1.272% to 0.999%, and RMSDP decreased from 1.412% to 1.172%. This shows that the effective information extraction of red ginseng terahertz spectrum by MCUVE is better than PCA, which is due to the fact that MCUVE can remove the noise signal without affecting the amount of effective information carried by the data.

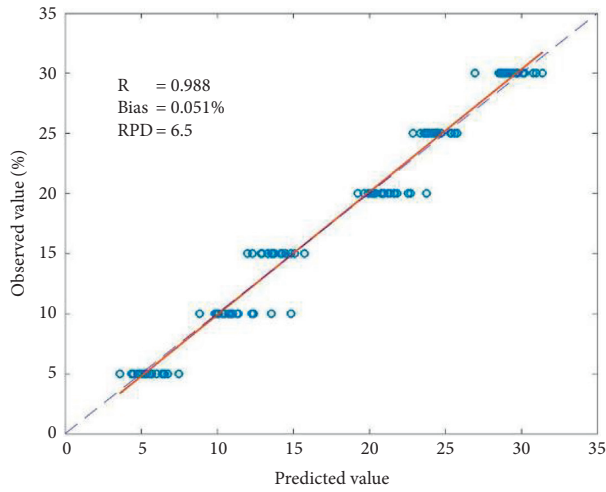
In order to visualize the evaluation of the model, observed (in the  $y$ -axis) vs. predicted (in the  $x$ -axis) (OP) regressions were used, and slope and intercept parameters were compared against the 1:1 line [35]. The evaluation graphs are shown in Figure 7. From the prediction of each

sucrose content, the prediction result of MCUVE-SVR is closer to the observed value, and the distribution is more concentrated, especially at the content of 15%, 20%, and 30%.

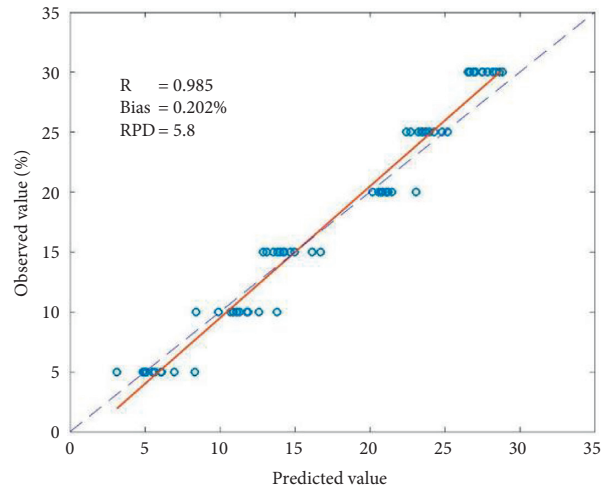
From the comparison of Bias, smaller absolute Bias values mean higher overall prediction accuracy. Arranged in descending order of absolute values of Bias, the models of the validation set are in the order PCA-PLSR, SVR, PLSR, MCUVE-PLSR, PCA-SVR, and MCUVE-SVR, while the models of the testing set are in the order PLSR, PCA-PLSR, MCUVE-PLSR, SVR, PCA-SVR, and MCUVE-SVR. It basically meets the expectation that SVR is more suitable than PLSR and MCUVE is more suitable than PCA in the detection application of this paper, and MCUVE-SVR obtained the best result. It is consistent with the evaluation results of  $R$  and RMSE.

From the comparison of RPD, the values of all models are greater than 5, which indicates that these models have reached a reliable level. The RPD values of the validation set and the testing set of MCUVE-SVR are 8.5 and 7.5, which are significantly larger than the corresponding values of other models, indicating that MCUVE-SVR has the highest reliability.

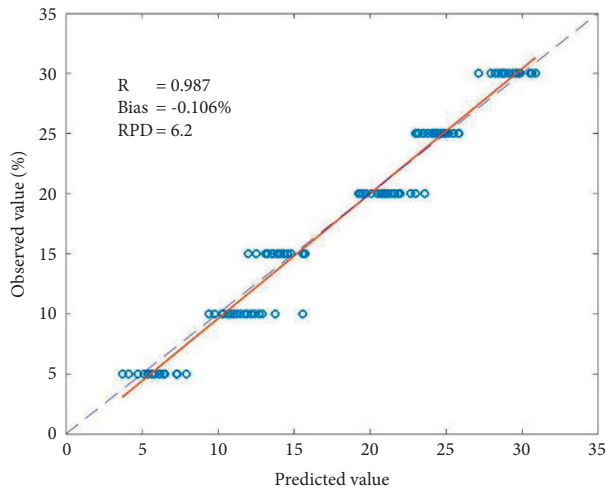
By comparing the values of  $R$ , RMSE, Bias, and RPD of each model, MCUVE-SVR is more suitable than the other five common models for the detection of adulterated sucrose in red ginseng.



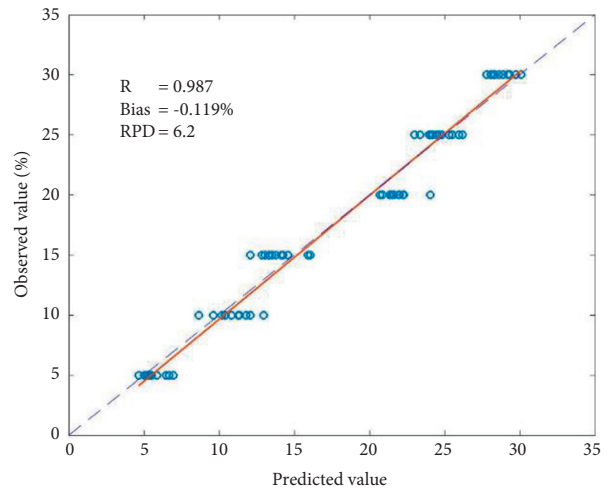
(a)



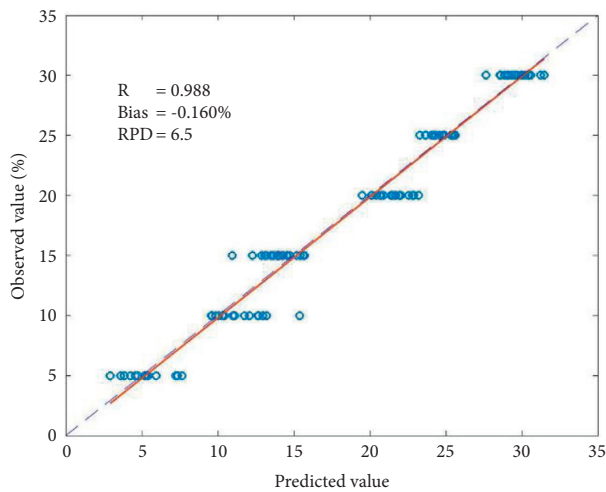
(b)



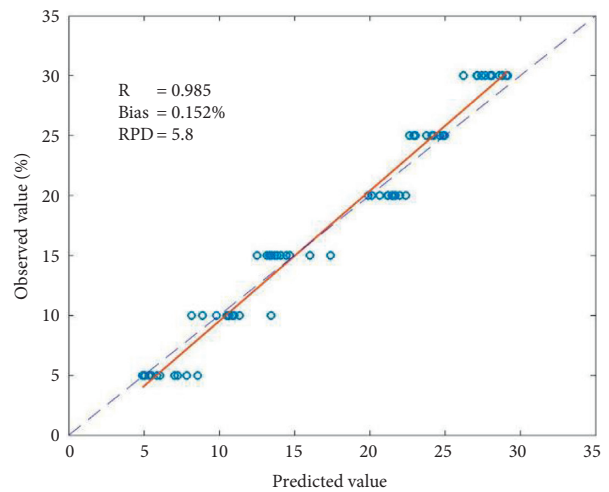
(c)



(d)



(e)



(f)

FIGURE 7: Continued.

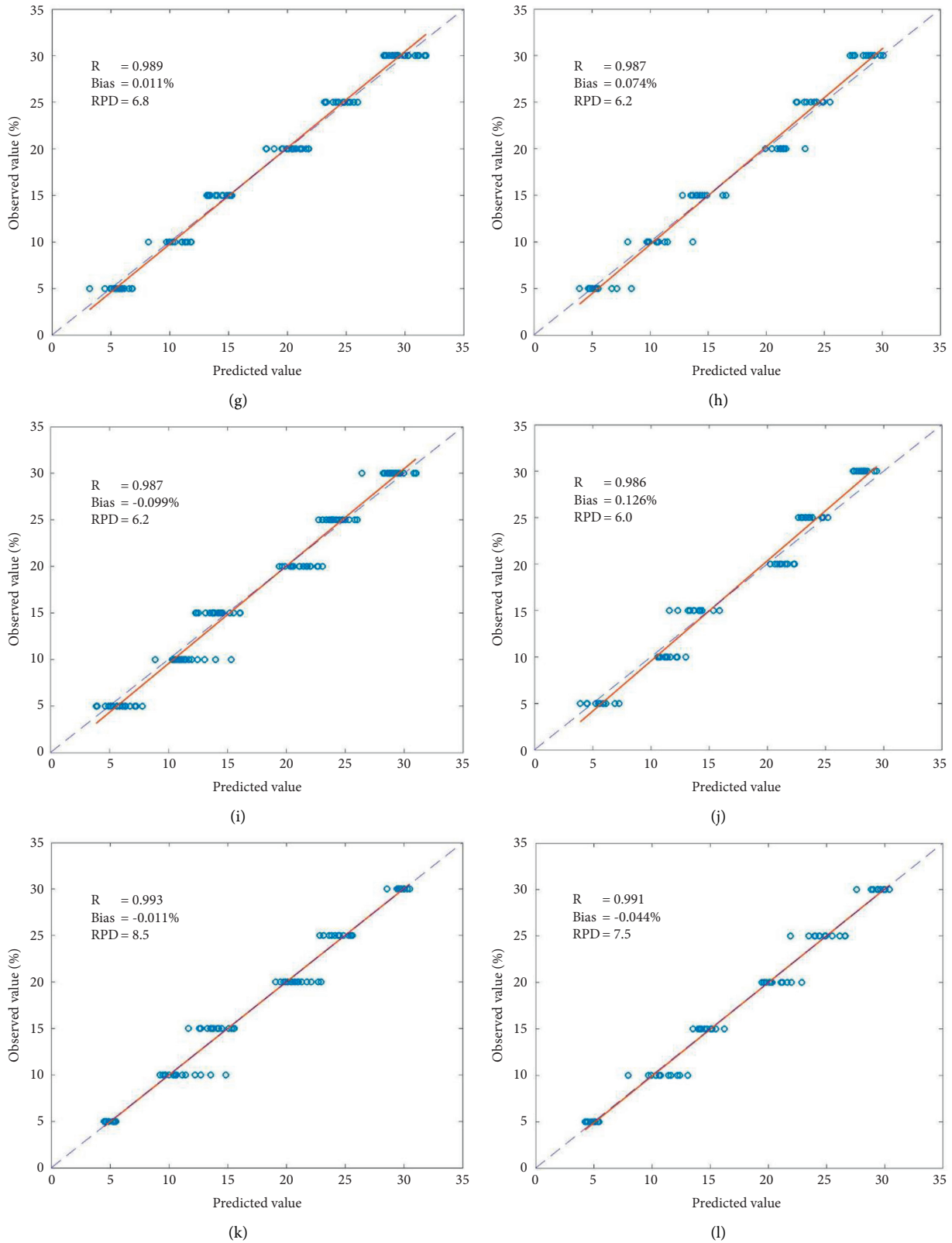


FIGURE 7: Correlation statistics between the observed values and predicted values: (a) validation set of PLSR, (b) testing set of PLSR, (c) validation set of SVR, (d) testing set of SVR, (e) validation set of PCA-PLSR, (f) testing set of PCA-PLSR, (g) validation set of PCA-SVR, (h) testing set of PCA-SVR, (i) validation set of MCVUE-PLSR, (j) testing set of MCVUE-PLSR, (k) validation set of MCVUE-SVR, and (l) testing set of MCVUE-SVR.



## 5. Conclusions

In order to improve the measurement accuracy of the sucrose content incorporated into red ginseng, we studied the quantitative analysis model of terahertz spectral characteristics. The experimental results show that the quantitative analysis model of sucrose based on terahertz spectroscopy can be more accurate by combining it with an effective information extraction algorithm. MCUVE is an effective information extraction algorithm commonly used in PLSR, which can improve the prediction accuracy of PLSR. In this paper, we try to use MCUVE in an SVR model and get more accurate prediction results. The RPD values of the validation set and the testing set of MCUVE-SVR reached 8.5 and 7.5, indicating that the model is ideal and better than other compared models. Therefore, MCUVE-SVR is a suitable quantitative regression model for the detection of adulterated sucrose in red ginseng. This paper has reference value and great meaning for the content detection of components in food and drug safety.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Acknowledgments

This work was supported in part by Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (Nos. YQ19103 and YQ20207), by Middle-aged and Young Teachers' Basic Ability Promotion Project of Guangxi (2020KY05024 and 2021KY0798), by National Natural Science Foundation of China (62063003), and by Guangxi Postdoctoral Special Fund.

## References

- [1] J. H. Heo, S. T. Lee, K. Chu et al., "An open-label trial of Korean red ginseng as an adjuvant treatment for cognitive impairment in patients with alzheimer's disease," *European Journal of Neurology*, vol. 15, no. 8, pp. 865–868, 2008.
- [2] J. Sun, C. Jiao, Y. Ma, J. Chen, W. Wu, and S. Liu, "Anti-ageing effect of red ginseng revealed by urinary metabolomics using RRLC-Q-TOF-MS," *Phytochemical analysis*, vol. 29, no. 4, pp. 387–397, 2018.
- [3] C. E. Hong and S. Y. Lyu, "Anti-inflammatory and anti-oxidative effects of Korean red ginseng extract in human keratinocytes," *Immune Network*, vol. 11, no. 1, pp. 42–49, 2011.
- [4] X. Xu, W. Qu, Z. Jia et al., "Effect of cultivation ages on anti-inflammatory activity of a new type of red ginseng," *Bio-medicine & Pharmacotherapy*, vol. 136, 2021.
- [5] S.-S. Zhou, K. Ka-Wai Auyeung, K.-M. Yip et al., "Stronger anti-obesity effect of white ginseng over red ginseng and the potential mechanisms involving chemically structural/compositional specificity to gut microbiota," *Phytomedicine*, vol. 74, Article ID 152761, 2018.
- [6] S. O. Suh, M. Kroh, N. R. Kim, Y. G. Joh, and M. Y. Cho, "Effects of red ginseng upon postoperative immunity and survival in patients with stage III gastric cancer," *The American Journal of Chinese Medicine*, vol. 30, no. 4, pp. 483–494, 2002.
- [7] X. F. Xu, Y. Gao, S. Y. Xu et al., "Remarkable impact of steam temperature on ginsenosides transformation from fresh ginseng to red ginseng," *Journal of Ginseng Research*, vol. 42, no. 3, pp. 277–287, 2018.
- [8] X. Ying, J. Li, W. Kong, D. Li, and L. Liu, "Rapid analysis of multi-components of ginseng radix et rhizoma rubra based on near infrared spectroscopy," *Chinese Journal of Modern Applied Pharmacy*, vol. 34, no. 10, pp. 1377–1384, 2017.
- [9] D. Carina, R. G. Junqueira, and V. Scheilla, "Interlaboratory validation of modified classical qualitative methods for detection of adulterants in milk: starch, chloride, and sucrose," *Food Analytical Methods*, vol. 9, no. 9, pp. 2509–2520, 2016.
- [10] J. A. Campbell, R. W. Hansen, and J. R. Wilson, "Cost-effective colorimetric microtitre plate enzymatic assays for sucrose, glucose and fructose in sugarcane tissue extracts," *Journal of the Science of Food and Agriculture*, vol. 79, no. 2, pp. 232–236, 1999.
- [11] L. Zhang, S. Wang, B. Qu, H. Chi, Y. Quan, and X. Wu, "Efficient separation determination of protopanaxatriol ginsenosides rg1, re, rf, rh1, rg2 by HPLC," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 170, pp. 48–53, 2019.
- [12] W. Chen, D. He, and Y. Shao, "Determination of AF and AFG in red ginseng by high performance liquid chromatography with evaporative light scattering detector(HPLC-ELSD)," *Animal Husbandry and Feed Science*, vol. 11, no. Z2, pp. 9–12, 2019.
- [13] J. Yeon Choi, J. Ho Hong, Y. Mi Dang et al., "Identification markers of adulteration in Korean red ginseng (panax ginseng) products using high-performance liquid chromatography (HPLC) and liquid chromatography–mass spectrometry (LC-MS)," *Analytical Letters*, vol. 51, no. 16, pp. 2586–2599, 2018.
- [14] D. Yoon, W. C. Shin, Y. S. Lee, S. Kim, and D. Y. Lee, "A comparative study on processed panax ginseng products using hr-mas nmr-based metabolomics," *Molecules*, vol. 25, no. 6, 2020.
- [15] Y. Tian, Y. Lu, J. Xie et al., "Rapid determination of ginsenoside rg1, re and rb1 in ginseng samples by capillary electrophoresis," *Analytical Methods*, vol. 1, no. 3, 2009.
- [16] V. A. Trofimov and S. A. Varentsova, "Detection and identification of drugs under real conditions by using noisy terahertz broadband pulse," *Applied Optics*, vol. 55, no. 33, pp. 9605–9618, 2016.
- [17] B. Qin, Z. Li, Z. Luo, H. Zhang, and Y. Li, "Feasibility of terahertz time-domain spectroscopy to detect carbendazim mixtures wrapped in paper," *Journal of Spectroscopy*, vol. 2017, Article ID 6302868, 8 pages, 2017.
- [18] T. Kozai, T. Kaneko, N. Hirai, and Y. Ohki, "Identification of antioxidants in polymeric insulating materials by terahertz absorption spectroscopy," *Polymer Degradation and Stability*, vol. 147, pp. 284–290, 2017.
- [19] M. Grognot and G. Gallot, "Quantitative measurement of permeabilization of living cells by terahertz attenuated total reflection," *Applied Physics Letters*, vol. 18, no. 10, 2015.
- [20] L. Wang, X. Wu, Y. Peng et al., "Quantitative analysis of homocysteine in liquid by terahertz spectroscopy," *Biomedical Optics Express*, vol. 11, no. 5, pp. 2570–2577, 2020.

- [21] W. Liu, P. Zhao, Y. Shi, C. Liu, and L. Zheng, "Rapid determination of peroxide value of peanut oils during storage based on terahertz spectroscopy," *Food Analytical Methods*, vol. 14, no. 6, pp. 1269–1277, 2021.
- [22] Y. Jiang, H. Ge, F. Lian, Y. Zhang, and S. Xia, "Early detection of germinated wheat grains using terahertz image and chemometrics," *Scientific Reports*, vol. 6, no. 1, Article ID 21299, 2016.
- [23] Z. Tang, H. Deng, Q. Liu, J. Guo, and L. Shang, "Quantitative analysis of low-concentration  $\alpha$ -HMX based on terahertz spectroscopy," *Analytical Methods*, vol. 12, no. 24, 2020.
- [24] Y. Hao and X. D. Sun, "Variable selection methods combined with local linear embedding theory used for optimization of near infrared spectral quantitative models," *Spectroscopy and Spectral Analysis*, vol. 32, no. 12, p. 3208, 2012.
- [25] D. Dharma Silalahi, H. Midi, J. Arasan, M. Shafie Mustafa, and J.-P. Caliman, "Robust wavelength selection using filter-wrapper method and input scaling on near infrared spectral data," *Sensors*, vol. 20, no. 17, 2020.
- [26] A. D. Burnett, W. Fan, P. C. Upadhyaya et al., "Broadband terahertz time-domain spectroscopy of drugs-of-abuse and the use of principal component analysis," *The Analyst*, vol. 134, no. 8, pp. 1658–1668, 2009.
- [27] L. J. Ni, L. G. Zhang, J. Xie, and J. Q. Luo, "Pattern recognition of Chinese flue-cured tobaccos by an improved and simplified K-nearest neighbors classification algorithm on near infrared spectra," *Analytica Chimica Acta*, vol. 633, no. 1, pp. 43–50, 2009.
- [28] S. Hua Lu, B. Qiong Li, H. Lin Zhai, X. Zhang, and Z. Yong Zhang, "An effective approach to quantitative analysis of ternary amino acids in foxtail millet substrate based on terahertz spectroscopy," *Food Chemistry*, vol. 246, pp. 220–227, 2017.
- [29] S. Warnecke, J. X. Wu, A. Rinnan et al., "Quantifying crystalline  $\alpha$ -lactose monohydrate in amorphous lactose using terahertz time domain spectroscopy and near infrared spectroscopy," *Vibrational Spectroscopy*, vol. 102, pp. 39–46, 2019.
- [30] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, vol. 185, no. 1, pp. 1–17, 1986.
- [31] V. Centner and D. L. Massart, "Elimination of uninformative variables for multivariate calibration," *Analytical Chemistry*, vol. 68, no. 21, pp. 3851–3858, 1996.
- [32] W. Cai, Y. Li, and X. Shao, "A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 90, no. 2, pp. 188–194, 2008.
- [33] Y. Sun, J. Huang, L. Shan, S. Fan, Z. Zhu, and X. Liu, "Quantitative analysis of bisphenol analogue mixtures by terahertz spectroscopy using machine learning method," *Food Chemistry*, vol. 352, no. 3, 2021.
- [34] X. Wei, S. Li, S. Zhu et al., "Terahertz spectroscopy combined with data dimensionality reduction algorithms for quantitative analysis of protein content in soybeans," *Spectrochimica Acta Part A Molecular and Biomolecular Spectroscopy*, vol. 253, 2021.
- [35] G. Piñeiro, S. Perelman, J. P. Guerschman, and J. M. Paruelo, "How to evaluate models: observed vs. predicted or predicted vs. observed?" *Ecological Modelling*, vol. 216, no. 3-4, pp. 316–322, 2008.