applied sciences

MDPI

Article

A Rapid Nondestructive Detection Method for Liquor Quality Analysis Using NIR Spectroscopy and Pattern Recognition

Guiyu Zhang [1,2,3,4], Xianguo Tuo [1,2,3,*], Yingjie Peng [2,3], Xiaoping Li [2,3] and Tingting Pang [2,3]

[1] School of Information Engineering, Southwest University of Science and Technology, Mianyang 621000, China; gyz_118@163.com
[2] School of Automation & Information Engineering, Sichuan University of Science & Engineering, Yibin 644000, China
[3] Artificial Intelligence Key Laboratory of Sichuan Province, Yibin 644000, China
[4] Liquor Making Biological Technology and Application of Key Laboratory of Sichuan Province, Yibin 644000, China
* Correspondence: tuoxianguo@suse.edu.cn

Abstract: Liquor has a complex system with high dimensional components. The trace components in liquor are varied and have low content and complex coordination relationships. This study aimed to solve the problem of reliance on smell and taste. Based on the characteristics of near-infrared spectrum response to hydrogen-containing groups, qualitative analysis was carried out in combination with machine learning technology. Firstly, an iterative adaptive weighted penalized least squares algorithm with spectral peak discrimination was used for baseline correction to effectively retain useful information in the feature absorption peaks. Then, the convolution smoothing algorithm was used to filter the noise, and the spectral curve smoothness was adjusted using the convolution window width. The near-infrared spectrum has a high dimension. Monte Carlo random sampling combined with an improved competitive adaptive reweighting method was used to evaluate the importance of spectral sampling points. According to the importance coefficient, the dimension of the spectral data set was optimized by using an exponential attenuation function through an iterative operation, and the data set with the smallest root-mean-square error was taken as the characteristic spectrum. The nonlinear separability of characteristic spectra was further improved by kernel principal component analysis. Finally, a liquor quality recognition model based on principal component analysis was established by using the hierarchical multiclass support vector machine method. Our key findings revealed that the prediction accuracy of the model reached 96.87% when the number of principal components was 5–12, with more than 95% of the characteristic information retained. These results demonstrated that this rapid nondestructive testing method resolved the challenge posed by relying on subjective sensory evaluation for liquor analysis. The findings provide a reliable analytical approach for studying substances with high-dimensional component characteristics.

Keywords: near-infrared spectroscopy; rapid nondestructive detection; peak identification; characteristic extraction; nonlinear separability; multiple classification prediction model

check for updates

Citation: Zhang, G.; Tuo, X.; Peng, Y.; Li, X.; Pang, T. A Rapid Nondestructive Detection Method for Liquor Quality Analysis Using NIR Spectroscopy and Pattern Recognition. *Appl. Sci.* **2024**, *14*, 4392. https://doi.org/10.3390/app14114392

Academic Editor: Roberto Anedda

Received: 21 February 2024
Revised: 12 May 2024
Accepted: 14 May 2024
Published: 22 May 2024

1. Introduction

Chinese liquor is one of the six major distilled spirits in the world (baijiu, whiskey, vodka, gin, brandy, and rum) [1,2]. With a rich history spanning thousands of years, it has a unique solid-state brewing process. Its main characteristics include being an alcoholic beverage inoculated with natural microorganisms, using starch-rich grains as the raw material, and using Chinese koji as a saccharifying starter culture for solid-state saccharification and fermentation. It is produced by solid-state distillation, and then stored and blended [3,4]. Liquor has a complex system with multiple components, with ethanol and water accounting for 98%–99%, and thousands of aromas and flavor trace components accounting for the remaining 1%–2% [5,6]. In recent years, liquor has attracted increasing attention as a traditional alcoholic beverage and is loved by consumers. The coordination and comprehensive activity actions of trace components determine the unique flavor and taste of liquor [7,8]. However, problems such as shoddy production and adulteration often occur in liquor products, most of which are made into products with a similar style and alcohol content to regular brand liquors, which seriously affects consumer health and their experience with the product. Therefore, liquor quality identification is of great significance to food safety.

Many analytical methods have been used in liquor detection, such as spectrophotometry [9,10], gas chromatography [11,12], mass spectrometry [13], and high-performance liquid chromatography [14]. The above methods are mainly used for the quantitative detection of structure and coordination components in liquor. The content of structure components is generally higher than 2~3 mg/100 mL, and are the basis of liquor flavor. The composition of the structure components of different flavors and styles is different. Taking Luzhou-flavored liquor as an example, the structure components that have been studied include four esters and four acids (ethyl acetate, ethyl lactate, ethyl butyrate, and ethyl caproate; acetic acid, butyric acid, lactic acid, and caproic acid), main alcohols, ethyl valerate, ethyl formate, propionic acid, valerate, n-propanol, n-hexyl alcohol, and other substances [15–17]. The coordination components make both aroma and taste contributions, including acetaldehyde, acetal, acetic acid, lactic acid, caproic acid, and butyric acid [18–21]. Although these traditional detection and analysis methods have high accuracy, they involve a series of time-consuming experimental preparations, such as sampling, extraction, and chromatography, which cannot be applied in rapid nondestructive detection [22]. These detection methods are quantitative analyses of monomer compounds with high content, which cannot reflect the coordination and comprehensive effects of multiple components in liquor. Currently, it is impossible to determine the quality of liquor by quantitative analysis because the complex coordination relationship of multi-components has not been studied clearly [23,24]. In the quality control system of most liquor production enterprises, artificial sensory evaluation still occupies a dominant position. The inevitable defects of sensory evaluation, such as strong subjectivity, arbitrariness, and poor reproducibility, render quality control personnel unable to obtain an accurate quantitative evaluation result. Therefore, a rapid method for liquor quality identification is urgently needed. At present, near-infrared (NIR) spectroscopy is considered one of the most suitable methods for multi-component detection, with the advantages of a rapid, nondestructive identification of complex mixed objects and good reproducibility [25,26]. NIR spectroscopy is widely used in food safety, the quality control of agricultural products, and drug analysis. There has been a lot of research on NIR detection technology in the quantitative and qualitative analysis of wine composition and classification. In terms of quantitative analysis, Fernandez-Novales used visible light and NIR to analyze amino acids and total soluble solids in the ripening process of intact grape berries. Calibration models, cross-validation models, and prediction models were established by the partial least square method. The results showed that non-contact and nondestructive spectroscopy could be used as an alternative method to provide amino acid composition information on grapes [27]. Pascoa used the Savitzky–Golay preprocessing method to divide the spectrum into five bands. After an exhaustive search and the cross-verification of 31 combined bands, the optimal spectral region was determined. Then, partial least squares regression (PLSR) was used to establish a model for different parameters [28]. BaiX adopted Fourier near-infrared to solve the problem of band distribution and the quantitative prediction of rose oxides in Donglu wine. In order to improve the accuracy of model prediction, principal component analysis (PCA) was used to eliminate outliers. Synergical interval partial least square regression (Si-PLSR) was used to select eight synergical subintervals reflecting rose oxide content, which had good predictive performance [29]. In addition, near-infrared spectroscopy technology has also been studied in the quantitative analysis of wine alcohol and phenols [30–33]. In terms of qualitative analysis, Hu XZ used mid-infrared and NIR combined with chemometrics to identify wine origins. PCA, analog soft independent modeling (SIMCA), and discriminant analysis (DA) based on mid-infrared and NIR were used for modeling [34]. Nardi T used NIR to classify wines aged with different woods. PCA was first used to study the relationship between samples, and then orthogonal partial least squares (OPLS) regression was applied to establish a discrimination model [35]. Pan T. proposed a Bayesian classifier algorithm based on wavelength optimization and applied it to the recognition of wine brands by NIR, and the results showed that this method has potential in terms of food characterization, traceability, and the authenticity of the food substrate [36]. Ta N. traced the origin of wines based on an NIR-integrated machine learning method [37]. Li Z. identified and analyzed 730 Chinese liquors using various methods, and the prediction ability of the established PCA-LDA model reached 98.9% [38]. The above research indicates that NIR is an effective method for the quantitative and qualitative analysis of alcohol. NIR fingerprint detection and analysis techniques include preprocessing, feature extraction, and reduction to build a prediction model.

In the field of liquor detection, increasing numbers of researchers have applied NIR spectroscopy to quantitative analysis and the quality identification of liquor components. The quantitative analysis of liquor components based on NIR spectroscopy is relatively early in its development and mainly uses partial least squares (PLS) [39–41]. In the qualitative analysis of liquor, methods combining chemometrics and pattern recognition are mainly used, such as the identification of liquor flavor type [38], liquor brand [42], and liquor year [43]. These detection analysis and modeling methods mainly use principal component analysis (PCA), PLS, PLS discriminant analysis, support vector machine (SVM), linear discriminant analysis, and artificial neural network. NIR spectral data have high-dimensional characteristics and usually contain redundant information irrelevant or weakly related to sample characteristics, which affects the accuracy and generalization ability of the prediction model. Therefore, it is very important to mine the characteristic information of spectral data according to the detection purpose.

There are many methods for feature extraction and dimensionality reduction in NIR spectroscopy, such as uninformative variable elimination (UVE), competitive adaptive reweighted sampling (CARS), and the aforementioned PCA. UVE adds a random noise matrix to the original NIR spectral data, establishes a PLS regression model, uses the regression coefficient of the model to calculate the stability coefficient of each spectral wavelength, and screens the characteristic spectrum using a stability coefficient threshold [44]. CARS selects variables with larger absolute values of the regression coefficients in the PLS model using the adaptive reweighted sampling technique and removes variables with smaller absolute values to obtain characteristic spectra [45]. PCA maps high-dimensional data to a new space to obtain mutually orthogonal principal components to achieve dimensionality reduction [46].

Liquor quality detection includes the classification of the base liquor and the quality evaluation of the product liquor. This study uses the base liquor as the research object, which is the main raw material used to blend the product liquor. In the distillation process, the quality of the base liquor changes with distillation time. Therefore, in this study, three distilled base liquors, from the first, middle, and last stages, were collected. In the feature extraction method, CARS is a variable extraction algorithm based on Kaplan–Meier's survival of the fittest principle in Darwinian evolution. The CARS algorithm can effectively screen wavelength variables closely related to the properties of the measured substance, reduce collinearity between variables, and improve the accuracy of the model [47,48]. However, the spectrum still has a high dimension after component analysis (KPCA), which is conducive to improving the accuracy of the prediction model. KPCA is more suitable for nonlinear spectrum processing than PCA [49,50]. Here, a liquor quality model based on a multi-classification support vector machine (MSVM) was established using the characteristic spectrum after dimensionality reduction.

2. Materials and Methods

2.1. Liquor Samples

Base liquor samples were collected in 2022 at a distillery in Sichuan, China. Each layer of fermented grain in the cellar was evenly mixed, and then solid-state distillation was performed. Base liquor samples were identified by professional tasters and divided into three grades. Samples from different cellars were collected for each fermentation cycle, and samples of each grade were sampled at an equal time interval to the distillation time. Representative samples were retained after identification by professional tasters. Finally, 383 valid samples were obtained for each grade, totaling 1149 samples.

2.2. Near-Infrared Spectroscopy Measurement

A Fourier NIR spectrometer (Bruker, MATRIX-F, Ettlingen, Germany) with a transmission-type liquid probe was used. The spectrometer was preheated for 1 h before detection, and the detection parameters were as follows: ambient temperature, 20 °C; air relative humidity, <80%; spectral wavenumber range, 4000–12,500 cm^{-1}; and spectral resolution, 4 cm^{-1}. Each sample had 2203 sampling points.

A base liquor sample (5 mL) was placed in a sample bottle, and each sample was scanned 32 times to automatically obtain the average spectrum. The detection time for each sample was about 18 s. In order to avoid the influence of storage on the sample, the sample was examined by NIR immediately after identification by a professional taster.

2.3. Near-Infrared Spectroscopy Preprocessing Method

For nonlinear baseline correction, this study used the iterative adaptive weighted penalized least squares method, which is based on an iterative weighting strategy of the error. The weight of the sum of squared residuals between the fitting baseline and the original signal was adaptively adjusted during the iterative process, and the irregularly changing baseline was quickly and flexibly subtracted. A fitted baseline *z* was iteratively found that minimized the cost function *Q*, as follows:

$$Q^t = \sum_{i=1}^{m} \omega_i^t F^t + \lambda \sum_{j=2}^{m} \left| z_j^t - z_{j-1}^t \right|^2, \tag{1}$$

$$\omega_i^t = \begin{cases} 0 & x_i \geq z_i^{t-1} \\ e^{\frac{t(x_i - z_i^{t-1})}{|d^t|}} & x_i < z_i^{t-1} \end{cases}, \tag{2}$$

$$F^t = \sum_{i=1}^{m} (x_i - z_i^t)^2, \tag{3}$$

where F^t is the sum of squared residuals; ω is the weight of the sum of squared residuals; m is the number of spectral samples; λ is a penalty term, which controls the smoothness of the fitted spectral curve; and z^t is the value of the fitted baseline.

After baseline correction, the Savitzky–Golay (SG) convolution smoothing algorithm was used for further noise reduction preprocessing. The commonly used moving window averaging filtering method results in decreased spectral resolution and relatively serious signal distortion [51]. SG approximates the curve within the moving window using the weighted average method for polynomial least squares fitting. Each observation value is multiplied by a weight coefficient to reduce the influence of smoothing on useful information. The weight coefficients are obtained based on the least squares principle. The formula of the smoothed spectral data is as follows:

$$x_i' = \frac{1}{A} \sum_{j=-m}^{m} \omega_j x_{i+j}, \quad i = 1, \cdots, n - 2m \tag{4}$$

where A is the normalization constant; $2m$ is the smoothing window width; and ω is the weight coefficient.

Spectral data with a high signal-to-noise ratio were obtained using the abovementioned preprocessing method, and then the first derivative method was used to improve the resolution and sensitivity of the spectral data.

2.4. Near-Infrared Spectroscopy Characteristic Extraction Method

The CARS algorithm is suitable for dealing with complex and nonlinear optimization problems. It is a feature variable extraction method combining Monte Carlo and partial least squares regression (PLSR). The dependent variable in this study was a categorical attribute, and therefore the discriminant partial least squares regression (DPLSR) algorithm was used to improve the PLSR algorithm in CARS. The regression coefficient (b_i) of each wavelength was obtained using the DPLSR model. Then, the proportion of wavelength points to be retained for the next modeling round was calculated according to the exponential decay function (R_t), and the wavelength points with larger absolute value weights (ω_i) of regression coefficients were retained as a new subset to re-establish the DPLSR model. Monte Carlo cross-validation was used to select the DPLSR model with the minimum classification error, and the corresponding wavelength point of the NIR spectrum was the characteristic spectrum.

$$R_t = \mu e^{-k_t} \tag{5}$$

$$\mu = (\frac{n}{2})^{1/(N-1)} \tag{6}$$

$$k = \frac{\ln(n/2)}{N-1} \tag{7}$$

$$\omega_i = |b_i| / \sum_{i=1}^{m} |b_i| \tag{8}$$

where n is the number of wavelength points in the original NIR spectrum; m is the number of retained wavelength points in each iteration; and N is the number of Monte Carlo samples.

2.5. Near-Infrared Characteristic Spectrum Dimension Reduction Method

To improve the classification effect of the NIR spectrum and the generalization ability of the classification model, it was necessary to further reduce the dimension of the feature spectrum. KPCA is the best algorithm for this purpose. KPCA is a dimension reduction method for nonlinear data, which solves the problem of linear inseparability of the nonlinear data. Through a nonlinear mapping function $\phi(\cdot)$, the original space data $X = (x_1, x_2, \ldots, x_m)$ is projected into a high-dimensional feature space, and data processing based on PCA is performed in the high-dimensional feature space. The nonlinear mapping function is not easy to solve, and the inner product of the mapping function is expressed by the polynomial kernel function K.

$$X' = [\phi(x_1), \phi(x_2), \cdots, \phi(x_n)] \tag{9}$$

$$K(x, y) = < \phi(x), \phi(y) > \tag{10}$$

$$K(x, y) = (ax^T y + c)^d \tag{11}$$

where d is the exponent of the polynomial kernel, with $a > 0$ and $c \geq 0$.

2.6. Model Construction Method

SVM has the advantages of strong generalization and global optimization in solving nonlinear and high-dimensional data set pattern recognition. A typical SVM classification model builds a hyperplane for binary classification. For the liquor quality multi-classification model, this study used a hierarchical multi-classification support vector machine (H-MSVM) method. First, the k-medoids algorithm was used to obtain the center points of each cluster. Then, the average Euclidean distance di between the center point of one class cluster and the center point of other classes was calculated, with a smaller di indicating higher similarity.

$$d_i = \frac{\sum\limits_{j=1, j\neq i}^{n} dis(c_i, c_j)}{n} \tag{12}$$

where *dis(·)* is the Euclidean distance formula; c_i and c_j are the cluster centers; and n is the number of classes in the dataset.

The cluster with the smallest d_i and other clusters were used as the first layer of the H-MSVM model for binary classification. The sample of this study contained three classes. Therefore, the other two classes were used as the second layer for binary classification. After the classification order was determined, the SVM classifier was trained for each layer.

3. Results and Discussion

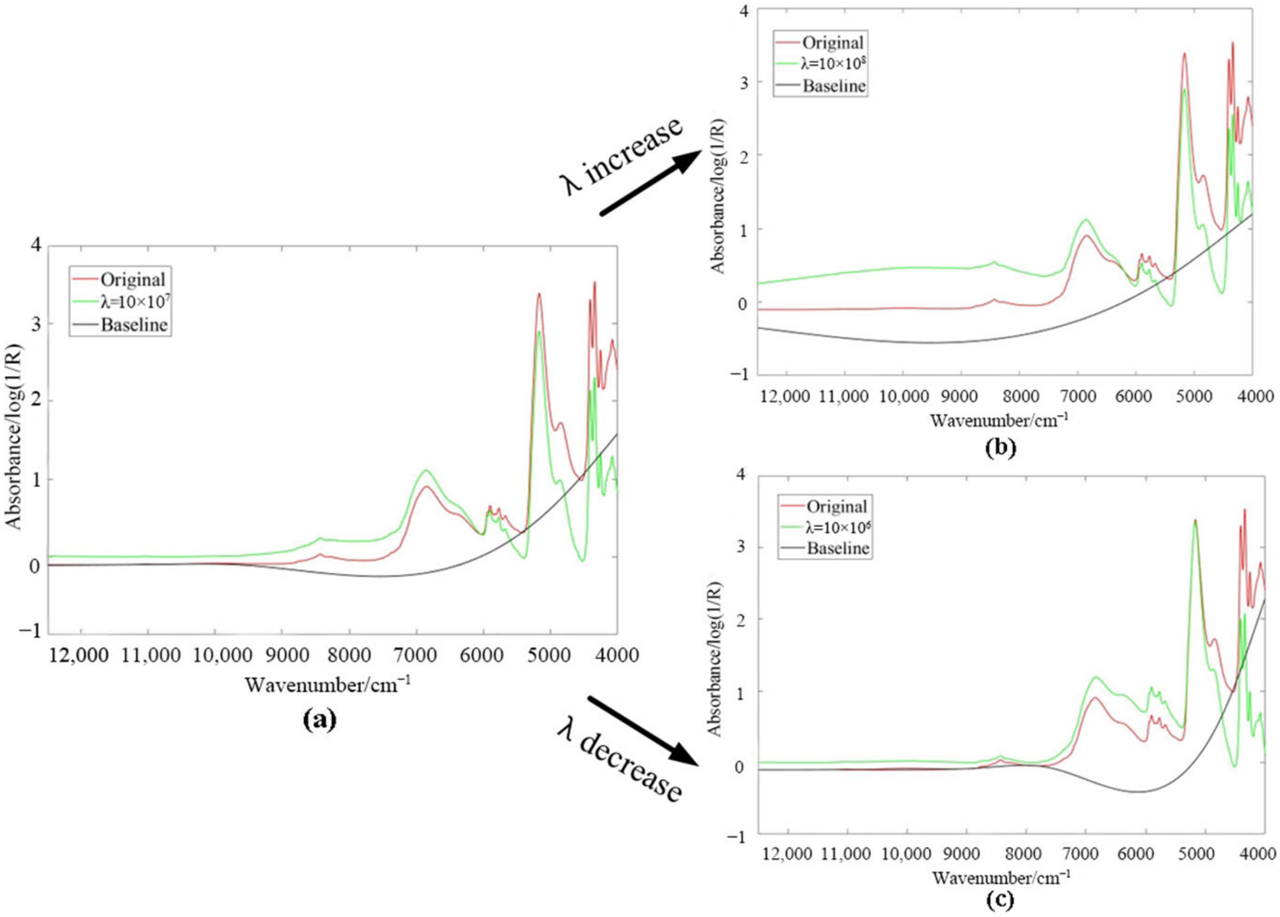
3.1. Near-Infrared Spectroscopy Preprocessing Results

The composition of NIR spectroscopy data is complex. In addition to characteristic variables, there is also baseline drift or background noise caused by the acquisition environment and equipment. Therefore, useless information must be filtered from the high-dimensional, non-zero variable complex data before modeling, to effectively mine the characteristic data of the spectrum. Because the base liquor is a transparent liquid, NIR detection utilizes the transmission principle of the spectrum. The commonly used baseline correction method for transmission spectra is the first derivative method [52,53] to increase the resolution of the spectral signal. However, the random noise in an NIR spectrum is generally a high-frequency signal, and the derivative method will amplify the noise signal. Therefore, effective baseline correction and noise filtering methods should be selected first to accurately extract the NIR feature spectra.

First, the adaptive iterative reweighted penalized least squares method was used for baseline correction. According to Formula (2), when the value of the original spectrum x^t was greater than the fitted baseline value, it was regarded as the spectral peak, and the weight was set to 0 to effectively retain the characteristics of the spectral data. Figure 1 shows the influence of different penalty terms on the smoothness of the fitted curve, in which the spectral peak features could be completely retained. When the penalty term was increased to $\lambda = 10 \times 10^8$ (Figure 1b), overfitting gradually occurred. The baseline correction effect was better when λ was set to 10×10^6 and 10×10^7, and the value of the penalty term was further determined according to the accuracy of the prediction model.

Then, the SG convolution smoothing algorithm was used for filtering, and the spectral resolution was enhanced using the first derivative method (Figure 2). Comparing the first derivative of the original spectral data in Figure 2A with the first derivative of the preprocessed spectral data in Figure 2B, it can be seen that the first derivative after preprocessing effectively eliminated the influence of high-frequency noise. The first derivative method can only remove part of the linear baseline and noise signal, but can also amplify the interference signal. Therefore, the baseline correction and filtering of NIR spectra are required before applying the first derivative. Finally, a multi-class SVM prediction model was established, and the optimal baseline correction and smoothing filter parameters were selected based on the results of the prediction model (Figure 3).

It can be seen from Figure 3A that the stability of the classification model was not high when the baseline correction penalty term $\lambda = 10 \times 10^6$. When the penalty term $\lambda = 10 \times 10^7$, as in Figure 3B, the prediction accuracy of the classification model was relatively stable. When the penalty term $\lambda = 10 \times 10^8$ (Figure 3C), as mentioned earlier, the baseline correction showed an overfitting phenomenon, resulting in low prediction accuracy of the classification model. Preprocessing parameters with a prediction accuracy greater than 86% were selected as the basis for further research (Table 1).



(a)

(b)

(c)

Figure 1. Effect of parameter λ on baseline correction of near-infrared spectroscopy. (**a**) Baseline correction penalty factor $\lambda = 10 \times 10^7$, (**b**) baseline correction penalty factor $\lambda = 10 \times 10^8$, and (**c**) baseline correction penalty factor $\lambda = 10 \times 10^6$.

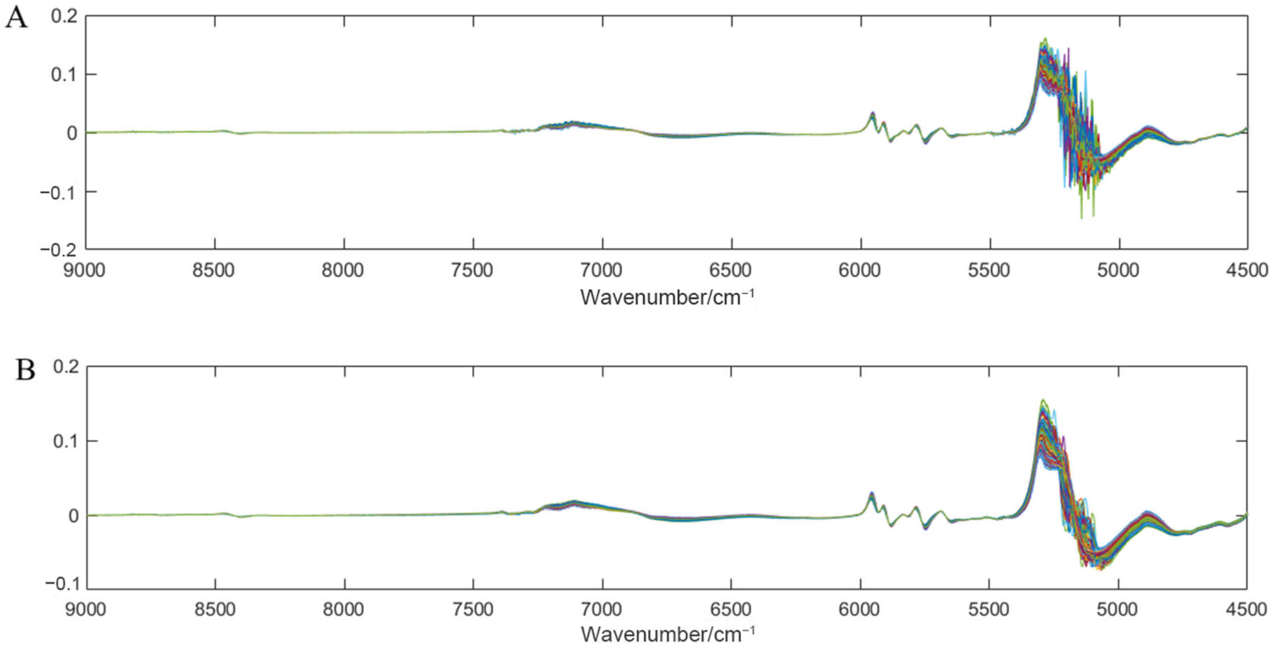


Figure 2. Preprocessed NIR spectra based on the first derivative. (**A**) Raw NIR spectra, and (**B**) baseline correction and filtered NIR spectra.

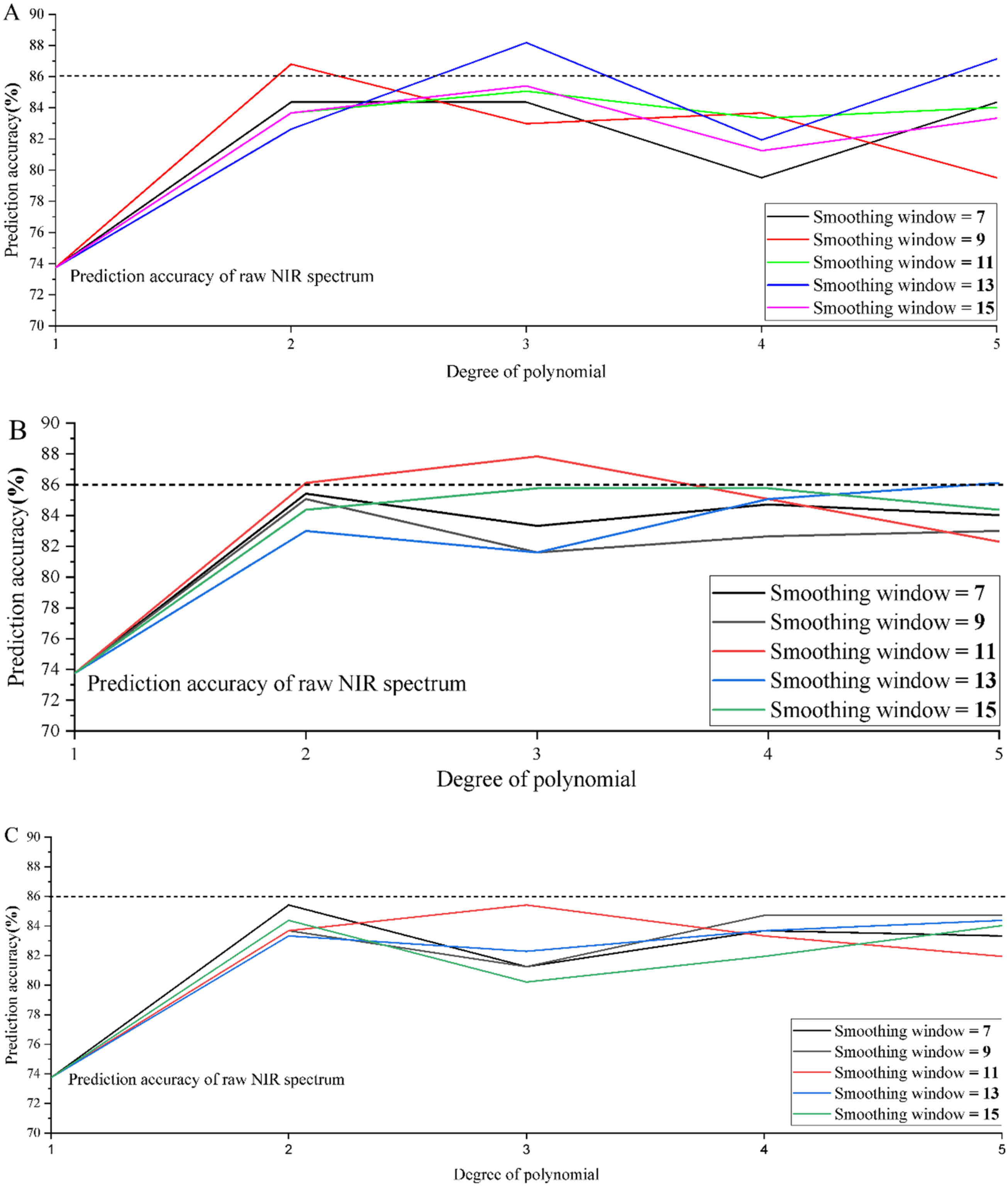


Figure 3. Comparison of the results of different preprocessing parameters based on a multi-classification support vector machine. (**A**) Baseline correction penalty factor λ = 10 × 10^6, (**B**) baseline correction penalty factor λ = 10 × 10^7, and (**C**) baseline correction penalty factor λ = 10 × 10^8.

Table 1. Results of different preprocessing parameters based on a multi-classification support vector machine.

Preprocessing Method	Baseline Correction Penalty Factor, λ	Smoothing Window	Degree of Polynomial	Prediction Accuracy (%)
1	10×10^6	9	2	86.81
2	10×10^6	13	3	88.19
3	10×10^6	13	5	87.15
4	10×10^7	11	2	86.11
5	10×10^7	11	3	87.85
6	10×10^7	13	5	86.11

3.2. Characteristic Wavelength Extraction

The differences in the liquor NIR spectra are very small, and therefore it is necessary to extract the characteristic wavelength that represents the liquor's characteristic information. The CARS method can effectively select wavelength variables closely related to the characteristics of the measured material, eliminate interference from redundant information, and improve the accuracy of the prediction model. Therefore, in this study, an improved CARS method was used to extract the characteristic wavelengths of the NIR spectrum. The second set of preprocessing parameters shown in Table 1 was used as an example (where the baseline correction penalty term $\lambda = 10 \times 10^6$, the smoothing window was 13, and the polynomial degree was 3). The number of Monte Carlo sampling iterations was set to 50, and the five-fold cross-validation method was used to extract 70 characteristic wavelengths. The results are shown in Figure 4.

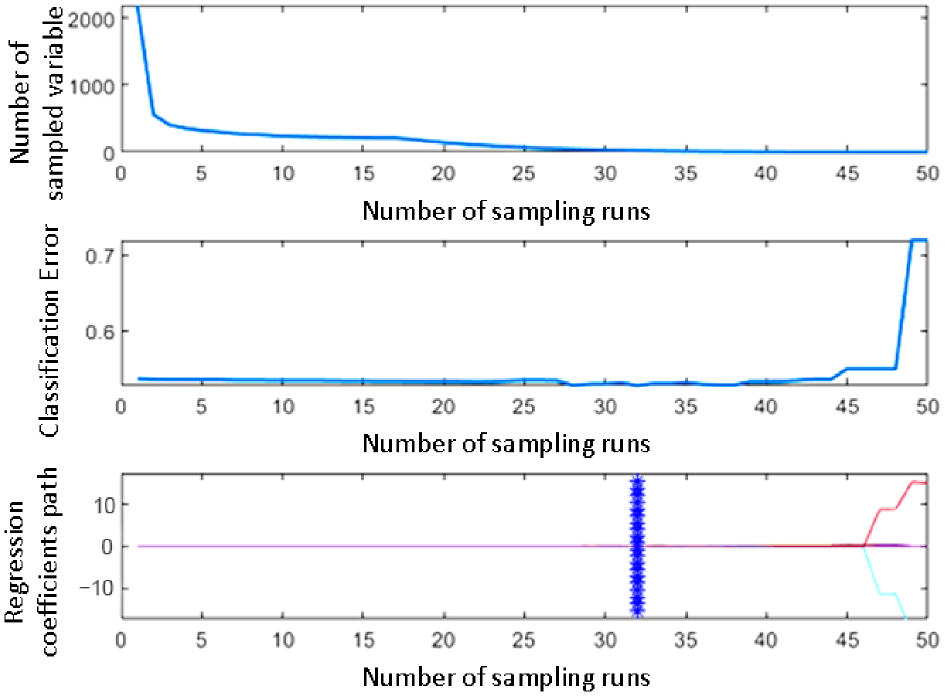


Figure 4. Characteristic wavelength extraction results.

The characteristic wavelength extraction results of the preprocessed data in Table 1 are shown in the "Characteristic wavelengths of improved CARS" section in Table 2. The results show that the NIR spectral feature extraction method of the base liquor based on improved CARS effectively reduces the dimension of the spectral data. The multi-class SVM prediction model was used to evaluate the feature wavelength extraction results, and the results showed that the prediction accuracy was improved.

Table 2. The results of characteristic wavelength extraction, dimension reduction, and model prediction.

Preprocessing Method	Characteristic Wavelengths of Improved CARS		Dimension Reduction of KPCA		Prediction Model Based on H-MSVM	
	Number of Characteristic Wavelengths	Prediction Accuracy (%)	Number of Principal Components	Principal Component Contribution Rate (%)	Prediction Accuracy (%)	Improved Prediction Accuracy (%)
1	52	89.34	9	95.21	92.51	3.17
2	70	92.72	5	95.33	96.87	8.68
3	69	90.59	7	95.39	93.66	3.07
4	50	89.80	8	95.15	93.05	3.25
5	52	90.24	8	95.09	92.26	2.02
6	49	90.18	12	95.02	91.58	1.40

3.3. Characteristic Wavelength Dimension Reduction

To reduce the complexity of the prediction model and improve prediction accuracy, it was necessary to further reduce the dimension based on the characteristic wavelength. The characteristic wavelength extraction results shown in Figure 4 were used for KPCA dimension reduction, and the results are shown in Figure 5.

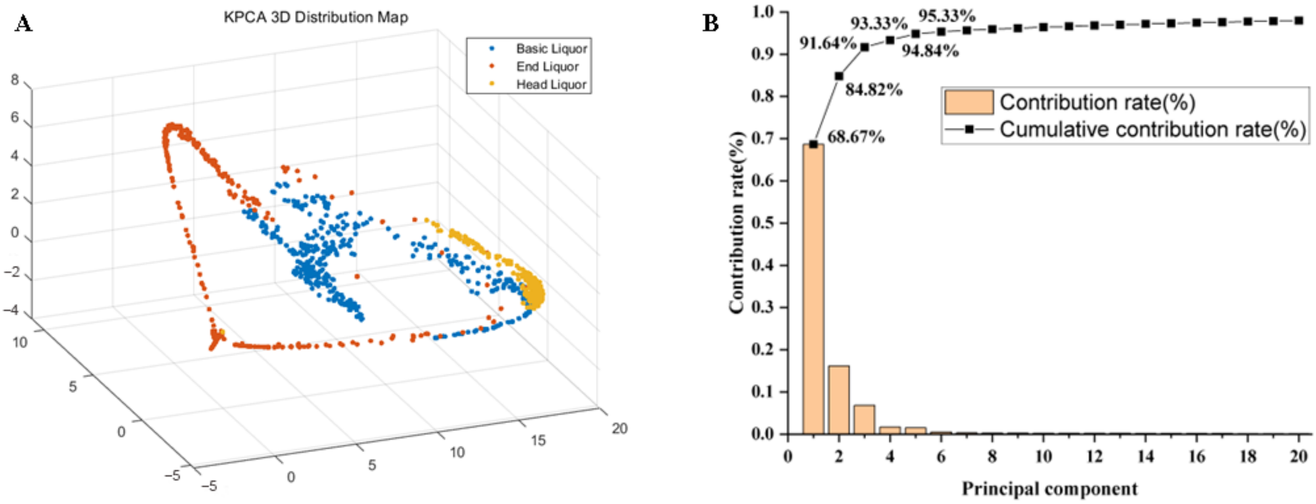


Figure 5. (**A**) The 3D distribution map of principal components, and (**B**) principal component contribution rate.

The KPCA dimensionality reduction method uses the Gaussian kernel function, and the 3D distribution map of the first three principal components with the highest contribution rate is shown in Figure 5A. The results show that KPCA improved the clustering effect of the samples. The contribution rates of the principal components and the cumulative contribution rates are shown in Figure 5B. To retain the main characteristic information of the original sample data, principal component selection requires that the cumulative contribution rate reaches more than 95%. The KPCA dimension reduction results for the characteristic wavelengths of the NIR spectrum are shown in the "Dimension reduction of KPCA" section of Table 2. The results show that the KPCA method effectively reduced the data dimension while retaining most of the features of the original sample.

3.4. Prediction of Liquor Quality

In this study, the SPXY dataset partitioning method was used to divide the samples into a training set and a test set, which accounted for three-quarters and one-quarter of the sample set, respectively. The method used near-infrared spectrum variables and classification label variables to calculate the distance between samples to characterize the sample distribution and effectively cover the vector space. By increasing the diversity and representativeness of sample division, the stability of the model can be improved.

The principal components obtained based on KPCA were used as input parameters for H-MVSM model learning. The accuracy of the model for liquor quality identification was verified through the test set, as shown in the "Prediction Model based on H-MSVM" section of Table 2.

The higher the number of characteristic wavelengths obtained based on the improved CARS, the fewer the number of principal components with the cumulative contribution rate of KPCA reaching more than 95%, and the higher the prediction accuracy of the model. Therefore, characteristic wavelength extraction did not indicate that lowering the dimension improved the model; instead, as much as possible the feature information in the original NIR spectrum should be retained. Through KPCA dimensionality reduction, the complexity of the model was reduced, and discrimination accuracy was improved. Using the second set of NIR spectroscopy preprocessing parameters as an example, when the number of characteristic wavelengths was 70, the accuracy of the model trained by the first five principal components improved by 8.68%, reaching a high of 96.87%.

4. Conclusions

In this study, NIR spectroscopy was used to rapidly identify liquor quality. Baseline correction and noise filter preprocessing were performed using a combination of adaptive iterative reweighted penalized least squares and SG convolutional smoothing. To ensure that more feature information was retained after the preprocessing of the NIR spectra, a multi-class SVM prediction model was used to evaluate different preprocessing parameters. Multiple sets of preprocessing parameters with the highest prediction accuracy were selected for feature extraction. The CARS method was used to extract feature wavelengths. The dependent variable in this study was the category attribute, and DPLSR was used to improve the CARS method. The KPCA method was used to further reduce the dimension of the feature wavelengths to improve the accuracy and generalization ability of the prediction model. The H-MSVM prediction model was built based on the principal components obtained by the KPCA method. Through the abovementioned methods, the accuracy of the prediction model was effectively improved, and the identification result of the best parameter combination reached more than 96%. The results showed that NIR is an effective tool for the nondestructive and rapid identification analysis of liquor quality, which can be applied to the on-line detection of liquor manufacturing and solve the problem of artificial sensory evaluation. At the same time, this method will provide an effective method for the qualitative analysis of substances containing high-dimensional components, and can be applied to the production of food, chemicals, and medical drugs.

Author Contributions: Conceptualization: X.T.; methodology and writing original draft: G.Z.; experiment and data acquisition: Y.P., X.L. and T.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Plan Project of Sichuan Provincial, grant number 2022YFS0554; the Liquor-Making Biological Technology and Application of the Key Laboratory of Sichuan Province Open Project, grant number NJ2022-06; the Scientific Research and Innovation Team Program of the Sichuan University of Science and Engineering, grant number SUSE652B005; Wuliangye and Sichuan University of Science and Engineering Industry—University —Research Project, grant number CXY2022ZR007; and the Science and Technology Achievement Transformation Project of Sichuan University of Science and Engineering, grant number HXJY01, China. The authors acknowledge this support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the components involved in the liquor.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Zhao, D.R.; Shi, D.M.; Sun, J.Y.; Sun, J.Y.; Li, A.J.; Sun, B.G.; Zhao, M.M.; Chen, F.; Sun, X.T.; Li, H.H.; et al. Characterization of key aroma compounds in Gujinggong Chinese Baijiu by gas chromatography-olfactometry, quantitative measurements, and sensory evaluation. *Food Res. Int.* **2018**, *105*, 616–627. [CrossRef]
2. Han, X.L.; Wang, D.L.; Zhang, W.J.; Jia, S.R. The production of the Chinese baijiu from sorghum and other cereals. *J. Inst. Brew.* **2017**, *123*, 600–604.
3. Wei, Q.; Lu, Z.M.; Chai, L.J.; Zhang, X.J.; Li, Q.; Wang, S.T.; Shen, C.H.; Shi, J.S.; Xu, Z.H. Cooperation within the microbial consortia of fermented grains and pit mud drives organic acid synthesis in strong-flavor Baijiu production. *Food Res. Int.* **2021**, *147*, 110449.
4. Wang, W.; Xu, Y.; Huang, H.; Pang, Z.; Fu, Z.; Niu, J.; Zhang, C.; Li, W.; Li, X.; Sun, B. Correlation between microbial communities and flavor compounds during the fifth and sixth rounds of sauce-flavor baijiu fermentation. *Food Res. Int.* **2021**, *150*, 110741. [CrossRef] [PubMed]
5. Wang, Q.; Liu, K.; Liu, L.; Zheng, J.; Shen, X.J. Correlation analysis between aroma components and microbial communities in Wuliangye-flavor raw liquor based on HS-SPME/LLME-GC-MS and PLFA. *Food Res. Int.* **2020**, *140*, 109995. [CrossRef] [PubMed]
6. Hong, J.X.; Zhao, D.R.; Sun, B.G. Research progress on the profile of trace components in baijiu. *Food Rev. Int.* **2021**, *39*, 1–27. [CrossRef]
7. Hong, J.X.; Tian, W.; Zhao, D. Research progress of trace components in sesame-aroma type of baijiu. *Food Res. Int.* **2020**, *137*, 109695. [CrossRef]
8. Fan, H.Y.; Fan, W.L.; Xu, Y. Characterization of key odorants in Chinese chixiang aroma-type liquor by gas chromatography-olfactometry, quantitative measurements, aroma recombination, and omission studies. *J. Agric. Food Chem.* **2015**, *63*, 3660–3668. [CrossRef] [PubMed]
9. Jamaludeen, A.; Neelamegam, P.; Rajendran, A. Determination of Calcium in Wine Using Reconfigurable PSoC Based Spectrophotometer. *J. Anal. Chem.* **2014**, *69*, 270–275. [CrossRef]
10. Wu, Z.; Xu, E.; Li, J.; Long, J.; Jiao, A.; Jin, Z.; Xu, X. Determination of Antioxidant Capacity of Chinese Rice Wine and Zhuyeqing Liquor Using Nanoparticle-Based Colorimetric Methods. *Food Anal. Methods* **2017**, *10*, 788–798. [CrossRef]
11. Wang, Z.; Wang, Y.; Zhu, T.T.; Wang, J.; Huang, M.Q.; Wei, J.W.; Ye, H.; Wu, J.H.; Zhang, J.L.; Meng, N. Characterization of the key odorants and their content variation in Niulanshan Baijiu with different storage years using flavor sensory omics analysis. *Food Chem.* **2022**, *376*, 131851. [CrossRef] [PubMed]
12. Yu, Y.; Chen, S.; Nie, Y.; Xu, Y. Optimization of an intra-oral solid-phase microextraction (SPME) combined with comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry (GC × GC-TOFMS) method for oral aroma compounds monitoring of Baijiu. *Food Chem.* **2022**, *385*, 132502. [CrossRef] [PubMed]
13. He, F.; Duan, J.W.; Zhao, J.W.; Li, H.H.; Sun, J.Y.; Huang, M.Q.; Sun, B.G. Different distillation stages Baijiu classification by temperature-programmed headspace-gas chromatography-ion mobility spectrometry and gas chromatography-olfactometry-mass spectrometry combined with chemometric strategies. *Food Chem.* **2021**, *365*, 130430. [CrossRef] [PubMed]
14. Wu, Z.; Qin, D.; Duan, J.W.; Li, H.H.; Sun, B.G. Characterization of benzenemethanethiol in sesame-flavour baijiu by high-performance liquid chromatography-mass spectrometry and sensory science. *Food Chem.* **2021**, *1*, 130345. [CrossRef]
15. Wang, J.S.; Chen, H.; Wu, Y.S.; Zhao, D.R. Uncover the flavor code of strong-aroma baijiu: Research progress on the revelation of aroma compounds in strong-aroma baijiu by means of modern separation technology and molecular sensory evaluation. *J. Food Compos. Anal.* **2022**, *109*, 104499. [CrossRef]
16. Xu, Y.Q.; Zhao, J.R.; Liu, X.; Zhang, C.S.; Sun, B.G. Flavor mystery of Chinese traditional fermented baijiu: The great contribution of ester compounds. *Food Chem.* **2022**, *369*, 130920. [CrossRef] [PubMed]
17. Wu, Y.S.; Hou, Y.X.; Chen, H.; Wang, J.S.; Zhang, C.S.; Zhao, Z.G.; Ao, R.; Huang, H.; Hong, J.X.; Zhao, D.R.; et al. "Key factor" for baijiu quality: Research progress on acid substances in baijiu. *Foods* **2022**, *11*, 2959. [CrossRef]
18. Dong, W.; Dai, X.R.; Jia, Y.T.; Ye, S.T.; Shen, C.H.; Liu, M.; Lin, F.; Sun, X.T.; Xiong, Y.F.; Deng, B. Association between Baijiu chemistry and taste change: Constituents, sensory properties, and analytical approaches. *Food Chem.* **2023**, *437*, 137826. [CrossRef] [PubMed]
19. Wei, Y.; Zou, W.; Shen, C.H. Basic flavor types and component characteristics of Chinese traditional liquors: A review. *J. Food Sci.* **2020**, *85*, 4096–4107. [CrossRef]
20. Liu, Y.B.; Qiao, Z.N.; Zhao, Z.J.; Wang, X.; Sun, X.Y.; Han, S.N.; Pan, C.M. Comprehensive evaluation of Luzhou-flavor liquor quality based on fuzzy mathematics and principal component analysis. *Food Sci. Nutr.* **2022**, *10*, 1780–1788. [CrossRef]

21. Wang, G.N.; Jing, S.; Wang, X.L.; Zheng, F.P.; Li, H.H.; Sun, B.; Li, Z.X. Evaluation of the perceptual interaction among ester odorants and nonvolatile organic acids in Baijiu by GC-MS, GC-O, odor threshold, and sensory analysis. *J. Agric. Food Chem.* **2022**, *70*, 13987–13995. [CrossRef] [PubMed]
22. Jia, W.; Fan, Z.; Du, A.; Li, Y.L.; Zhang, R.; Shi, Q.Y.; Shi, L.; Chu, X.G. Recent advances in Baijiu analysis by chromatography based technology—A review. *Food Chem.* **2020**, *324*, 126899. [CrossRef] [PubMed]
23. Fan, T.; Xie, Z.M.; Wei, J.P.; Zheng, J.; Huang, J.; Zhang, Q. Health risk exposure assessment of eight heavy metals in world distilled spirits. *China Brew.* **2023**, *42*, 228–233.
24. Song, D.D.; He, P.H.; Chen, J.; Yang, L. Difference of volatile substances in top six distilled spirits based on GC-MS. *China Brew.* **2020**, *39*, 190–195.
25. Sohn, S.; Pandian, S.; Oh, Y.J.; Zaukuu, J.L.Z.; Kang, H.J.; Ryu, T.H.; Cho, W.S.; Cho, Y.S.; Shin, E.K.; Cho, B.K. An Overview of Near Infrared Spectroscopy and Its Applications in the Detection of Genetically Modified Organisms. *Int. J. Mol. Sci.* **2021**, *22*, 9940. [CrossRef]
26. Luo, W.; Tian, P.; Fan, G.Z.; Dong, W.T.; Zhang, H.L.; Liu, X.M. Non-destructive determination of four tea polyphenols in fresh tea using visible and near-infrared spectroscopy. *Infrared Phys. Technol.* **2022**, *123*, 104037. [CrossRef]
27. Fernández-Novales, J.; Garde-Cerdán, T.; Tardáguila, J.; Gutierrez, G.G.; Pilar, P.A.E.; Paz, D.M. Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening. *Talanta* **2019**, *199*, 244–253. [CrossRef]
28. Páscoa, R.N.M.J.; Porto, P.A.L.S.; Cerdeira, A.L.; João, A.L. The application of near infrared spectroscopy to wine analysis: An innovative approach using lyophilization to remove water bands interference. *Talanta* **2020**, *214*, 120852. [CrossRef]
29. Bai, X.B.; Xu, Y.Q.; Chen, X.L.; Dai, B.X.; Tao, Y.S.; Xiong, X.L. Analysis of Near-Infrared Spectral Properties and Quantitative Detection of Rose Oxide in Wine. *Agronomy* **2023**, *13*, 1123. [CrossRef]
30. Anjos, O.; Caldeira, I.; Pedro, S.I. Screening of Different ageing technologies of wine spirit by application of near-infrared (NIR) spectroscopy and volatile quantification. *Processes* **2020**, *8*, 736. [CrossRef]
31. Rouxinol, M.I.; Martins, M.R.; Murta, G.C.; Barroso, J.M.; Rato, A.E. Quality assessment of red wine grapes through NIR spectroscopy. *Agronomy* **2022**, *12*, 637. [CrossRef]
32. Tzanova, M.; Atanassova, S.; Atanasov, V.; Grozeva, N. Content of polyphenolic compounds and antioxidant potential of some Bulgarian red grape varieties and red wines, determined by HPLC, UV, and NIR spectroscopy. *Agriculture* **2020**, *10*, 193. [CrossRef]
33. Anjos, O.; Caldeira, I.; Fernandes, T.A.; Pedro, S.I.; Vitória, C.; Alves, S.O.; Catarino, S.; Canas, S. PLS-R calibration models for wine spirit volatile phenols prediction by near-infrared spectroscopy. *Sensors* **2021**, *22*, 286. [CrossRef] [PubMed]
34. Hu, X.Z.; Liu, S.Q.; Li, X.H.; Wang, C.X.; Ni, X.L.; Liu, X.; Wang, Y.; Liu, Y.; Xu, C.H. Geographical origin traceability of Cabernet Sauvignon wines based on Infrared fingerprint technology combined with chemometrics. *Sci. Rep.* **2019**, *9*, 8256. [CrossRef] [PubMed]
35. Nardi, T.; Petrozziello, M.; Girotto, R.; Fugaro, M.; Mazzei, R.A.; Scuppa, S. Wine aging authentication through Near Infrared Spectroscopy: A feasibility study on chips and barrel aged wines. *OENO One* **2020**, *54*, 165–173. [CrossRef]
36. Pan, T.; Li, J.Q.; Fu, C.L.; Chang, N.L.; Chen, J.M. Visible and near-infrared spectroscopy combined with bayes classifier based on wavelength model optimization applied to wine multibrand identification. *Front. Nutr.* **2022**, *9*, 796463. [CrossRef] [PubMed]
37. Ta, N.; Wei, H.C.; Hu, Z.L.; Cao, X.H.; Xiao, M.X. Wine component tracing method based on near infrared spectrum fusion machine learning. *Front. Sustain. Food Syst.* **2023**, *7*, 1197508. [CrossRef]
38. Li, Z.; Wang, P.P.; Huang, C.C.; Shang, H.; Pan, S.Y.; Li, X.J. Application of Vis/NIR spectroscopy for Chinese liquor discrimination. *Food Anal. Methods* **2014**, *7*, 1337–1344. [CrossRef]
39. Shen, F.; Niu, X.Y.; Yang, D.T.; Ying, Y.B.; Li, B.B.; Zhu, G.Q.; Wu, J. Determination of amino acids in Chinese rice wine by fourier transform near-infrared spectroscopy. *J. Agric. Food Chem.* **2010**, *58*, 9809–9816. [CrossRef]
40. Ouyang, Q.; Zhao, J.; Chen, Q. Measurement of non-sugar solids content in Chinese rice wine using near infrared spectroscopy combined with an efficient characteristic variables selection algorithm. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2015**, *151*, 280–285. [CrossRef]
41. Han, S.H.; Zhang, W.W.; Li, X.; Li, P.Y.; Liu, J.X.; Luo, D.L.; Xu, B.C. Rapid determination of ethyl pentanoate in liquor using Fourier transform near-infrared spectroscopy coupled with chemometrics. *Spectrosc. Lett.* **2016**, *49*, 464–468. [CrossRef]
42. Yang, B.; Yao, L.; Tao, P. Near-Infrared Spectroscopy Combined with Partial Least Squares Discriminant Analysis Applied to Identification of Liquor Brands. *Engineering* **2018**, *9*, 181–189. [CrossRef]
43. Ning, Y.; Zhang, H.M.; Zhang, Q.; Zhang, X.R. Rapid identification and quantitative pit mud by near infrared Spectroscopy with chemometric. *Vib. Spectrosc.* **2020**, *110*, 103116. [CrossRef]
44. Zhang, Y.F.; Wang, Z.L.; Tian, X.; Yang, X.H.; Cai, Z.L.; Li, J.B. Online analysis of watercore apples by considering different speeds and orientations based on Vis/NIR full-transmittance spectroscopy. *Infrared Phys. Technol.* **2022**, *122*, 104090. [CrossRef]
45. Zhao, X.; Li, C.H.; Zhao, Z.L.; Wu, G.C.; Xia, L.Y.; Jiang, H.Z.; Wang, T.X.; Chu, X.; Liu, J. Generic models for rapid detection of vanillin and melamine adulterated in infant formulas from diverse brands based on near-infrared hyperspectral imaging. *Infrared Phys. Technol.* **2021**, *116*, 103745. [CrossRef]
46. Meléndez, J.; Guarnizo, G. Fast Quantification of Air Pollutants by Mid-Infrared Hyperspectral Imaging and Principal Component Analysis. *Sensors* **2021**, *21*, 2092. [CrossRef] [PubMed]

47. Pang, L.; Chen, H.; Yin, L.; Cheng, J.; Jin, J.; Zhao, H.; Liu, Z.; Dong, L.; Yu, H.; Lu, X. Rapid fatty acids detection of vegetable oils by Raman spectroscopy based on competitive adaptive reweighted sampling coupled with support vector regression. *Food Qual. Saf.* **2022**, *4*, 545–554. [CrossRef]
48. Xia, Z.Z.; Yang, J.; Wang, J.; Wang, S.P.; Liu, Y. Optimizing Rice Near-Infrared Models Using Fractional Order Savitzky–Golay Derivation (FOSGD) Combined with Competitive Adaptive Reweighted Sampling (CARS). *Appl. Spectrosc.* **2020**, *74*, 417–426. [CrossRef]
49. Tang, R.N.; Luo, X.C.; Li, C.; Zhong, S.X. A study on nitrogen concentration detection model of rubber leaf based on spatial-spectral information with NIR hyperspectral data. *Infrared Phys. Technol.* **2022**, *122*, 104094. [CrossRef]
50. Zheng, Y.S.; Chen, W.M.; Zhang, Y.N.; Bai, D.Y. Prediction of the Remaining Useful Life of a Switch Machine, Based on Multi-Source Data. *Sustainability* **2022**, *14*, 14517. [CrossRef]
51. Zhang, W.L.; Tian, F.C.; Song, A.; Hu, Y.W. Research on an Optical E-nose Denoising Method Based on LSSVM. *Optik* **2018**, *168*, 118–126. [CrossRef]
52. Liu, Y.; Sun, L.; Li, Y.; Zhang, D.; Song, Z.; Zhang, C.; Pan, X.; Li, J.; Wang, Z. Quality evaluation of fried soybean oil base on near infrared spectroscopy. *J. Food Process Eng.* **2018**, *41*, e12887. [CrossRef]
53. Zhang, X.; Gao, Z.; Yang, Y.; Pan, S.; Yin, J.; Yu, X. Rapid identification of the storage age of dried tangerine peel using a hand-held near infrared spectrometer and machine learning. *J. Near Infrared Spectrosc.* **2022**, *30*, 31–39. [CrossRef]