Hindawi
Journal of Food Quality
Volume 2021, Article ID 5527826, 14 pages
https://doi.org/10.1155/2021/5527826

Research Article

Age Discrimination of Chinese Baijiu Based on Midinfrared Spectroscopy and Chemometrics

Shiqi Hu[1,2] **and Le Wang**[3,4]

[1]*College of Food Science and Technology, Nanjing Agricultural University, Nanjing 210095, China*
[2]*Key Laboratory of Meat Processing and Quality ControlKey Laboratory of Meat Processing and Quality Control, MOE, Key Laboratory of Meat Processing, MOA, Jiangsu Synergetic Innovation Center of Meat Processing and Quality Control, Nanjing 210095, China*
[3]*College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*
[4]*Key Laboratory of Navigation, Control and Health-Management Technologies of Advanced Aerocraft, Ministry of Industry and Information Technology, Nanjing 211106, China*

Correspondence should be addressed to Le Wang; wanglemaths@nuaa.edu.cn

Received 26 February 2021; Revised 7 August 2021; Accepted 21 August 2021; Published 1 September 2021

Academic Editor: Walid Elfalleh

Baijiu is a traditional and popular Chinese liquor which is affected by the storage time. The longer the storage time of Baijiu is, the better its quality is. In this paper, the raw and mellow Baijiu samples from different storage time are discriminated accurately throughout midinfrared (MIR) spectroscopy and chemometrics. Firstly, changing regularities of the substances in Chinese Baijiu are discussed by gas chromatography-mass spectrometry (GC-MS) during the aging process. Then, infrared spectrums of Baijiu samples are processed by smoothing, multivariate baseline correction, and the first and second derivative processing, but no significant variation can be observed. Next, the spectral date pretreatment methods are constructively introduced, and principal component analysis (PCA) and discriminant analysis (DA) are developed for data analyses. The results show that the accuracy rates of samples by the DA method in calibration and validation sets are 91.7% and 100%, respectively. Consequently, an identification model based on support vector machine (SVM) and PCA is established combined with the grid search strategy and cross-validation methods to discriminate the age of Chinese Baijiu validly, where 100% classification accuracy rate is obtained in both training and test sets.

1. Introduction

Chinese Baijiu is one of the six distilled spirits in the world, and it is the most traditional and popular alcoholic drink with a history of more than 5000 years in China [1–3]. In the past three years, although Baijiu was suffering declining annual sales because of the impact of COVID-19, it has a huge consumer market. In 2020, the annual production of Baijiu reached 7.407 million hectoliters [4]. Therefore, the investigation of Chinese Baijiu in recent decades has attracted more and more interest. However, Chinese Baijiu is a transparent and extremely complex mixture. The most contents of Baijiu are water and alcohol [5], and Baijiu contains more than 300 organic compounds such as ethyl acetate, acetic acid, ethyl butyrate, and ethyl hexanoate and [6, 7], which only take up less than 3% volume fraction of it. It is widely known that these organic compounds determine the quality or flavor of Baijiu.

Flavor is the most important grading standard for Chinese Baijiu. In modern Baijiu industry, the aging process is usually employed to improve the flavor and quality of Chinese Baijiu. In other words, the age of Baijiu is the most important factor affecting flavor [8, 9], where the wine age is the storage years of Chines Baijiu in specific containers. That is because a series of slow physical and chemical reactions occurred during the extension of storage time. Some low-boiling impurities volatilize naturally such as sulfides, irritative aldehydes, and so on, which reduces the unpleasant

bitter taste and astringency. Meanwhile, due to the reinforced association between alcohol and water molecules and the volatilization of ethanol, the stimulation from alcohol has weakened compared with the high-proof raw Baijiu. In this case, more than 300 organic compounds can reach equilibrium, which forms more harmonious and coordinated taste and tends to achieve optimal quality and increasingly prominent fragrance [10–12]. Consequently, the liquor age is often used to evaluate the quality of Chinese Baijiu [13]. However, since better economic benefits can be obtained by prolonging the storage time, there exist many unacceptable phenomenon in the market, such as cutting corners in the production and treatment process, false reporting of the age of Chinese Baijiu. These cases disrupt the Baijiu market and seriously affect the reputation of Chinese Baijiu. Therefore, it is urgent to design a method to quickly and accurately detect the age of Chinese Baijiu and avoid the aforementioned problems [14, 15].

In recent decades, several technologies, used for the age detection and quality identification, have been proposed. The technologies mainly focus on the chromatography and spectrum, such as gas chromatography [16], gas chromatography-mass spectrometry (GC-MS) [17, 18], high-performance liquid chromatography [19], near-infrared spectroscopy [20], atomic absorption spectroscopy [21], visible-ultraviolet spectroscopy [22], and fluorescence spectroscopy [23]. In this paper, we adopt GC-MS and midinfrared (MID) spectroscopy technologies to classify the age of Chinese Baijiu. GC-MS is extensively applied in the field of spirit ingredient detection and accurate qualitative and quantitative analysis [17]. In [18], GC-MS, combined with an electronic nose system, was utilized to characterize the volatile aroma compounds in the Chinese Baijiu and distinguish the difference between different liquor ages.

MID spectroscopy is the absorption spectrum of material in the wavelength range of $2.5 \sim 25\,\mu$m. The information recorded in the MID spectrum is the fundamental absorption region of hydrogen-containing groups such as -CH, -NH, and -OH [24]. In the production of Chinese Baijiu, MID spectroscopy has developed into an effective approach for quantitative and qualitative analysis. In [25, 26], the aroma component detection has been performed, and the quantitative models for routine parameters in the spirit have been established. In [27, 28], the Baijiu samples from different geographical origins were classified accurately to realize the purpose of optimizing the brewing processing. In [29], MID is utilized to identify the authenticity of Chinese Baijiu for protecting the interests of consumers. The work in [30, 31] demonstrates the application of MIR spectroscopy in the classification of mellow wine. Nevertheless, neither of these studies the applications on the aging of Chinese Baijiu.

In addition, many intelligent models have been widely used in rapid detection of Chinese Baijiu due to their advantages in multivariate nonlinear modeling establishment. They are represented by principal component analysis (PCA) [32], artificial neural networks [33–35], and support vector machine (SVM) [36, 37]. In particular, SVM [38] is a learning method and first proposed by Cortes and Vapnik in 1995. It is based on statistical learning theory and structural risk minimization criterion. Meanwhile, it has a great superiority in solving the nonlinear and high-dimensional pattern recognition problems and other machine learning problems such as function fitting [39]. Therefore, the SVM model has already been widely employed in the food classification problems [40, 41]. In this paper, SVM is adopted to construct the classification model to realize the age discrimination of Chinese Baijiu.

In summary, it is important to investigate wine age discrimination of Chinese Baijiu based on midinfrared spectroscopy and chemometrics. In this paper, the aging mechanism of Baijiu is studied and a qualitative model is established to distinguish it from aging time (raw spirit, 1, 3, and 5 years old) according to infrared spectrum characteristics. Meanwhile, the impacts on the results of different spectral preprocessing methods, composing of the principal component analysis (PCA), discriminant analysis (DA), and SVM, are evaluated. The major contributions of this paper are summarized as follows:

(i) Based on near-infrared spectroscopy technology, a qualitative analysis method is developed to be able to quickly and nondestructively evaluate the age of Chinese Baijiu.

(ii) For the spectral data of Chinese Baijiu, PCA technology is proposed to extract the main data and exclude outliers to provide optimal variables for subsequent analysis. Simultaneously, PCA and DA are employed to establish the analysis model.

(iii) Furthermore, considering the limited number of the Baijiu samples, the grid search strategy and cross-validation methods are used to dynamically adjust the parameters of the SVM during the training process of the SVM classification model, which improves the accuracy of the SVM model.

This paper is organized as follows: materials and methods are listed in Section 2. The statistical analysis including GC-MS results, infrared spectrum data analysis, and DA model classification results are presented in Section 3. The constructing of SVM classification results is presented in Section 4. Section 5 gives the final conclusion and future work.

2. Materials and Methods

2.1. Experimental Material. Eighty Baijiu samples are provided by the Yanfeng Winery in Hunan, and the samples are selected from different workshops, vessels, and production dates. Luzhou-flavor Baijiu, whose alcohol content is 60% (V/V), is a typical fragrance type of Chinese Baijiu. Therefore, Luzhou-flavor Baijiu is selected in this paper. All samples are separated into four groups on the basis of storage time: 0, 1, 3, and 5 years. In total, 80 samples are collected and analyzed (20 samples of each group). Three-fourth of the samples are selected randomly for training the SVM model, namely, the training set. The remaining part is utilized to test the classification performance of the SVM model, namely, the test set. Furthermore, the training

sample set consists of 60 samples and the test sample set is composed of 20 samples. The distribution of Baijiu samples is listed in Table 1 in detail.

2.2. Determination of Volatile Aroma Components.

Chromatographic conditions: chromatographic column hp- 5 ms (30 m × 250 μm × 0.25 μm)

Front sample port temperature: 250°C

Carrier gas (helium) flow rate: 1 mL/min

Pressure: 2.4 kPa

Injection volume: 1 μL

Split ratio: 2: 1

Heating program: initial temperature 35°C for 5 min; 20°C/min to 230°C, for 2 min

Mass spectrometry conditions: EI ion source

Electron energy: 70 eV

Ion source temperature: 230°C

Quadrupole temperature: 150°C

Solvent delay: 2 min

Mass scanning range: $M/Z35 \sim 500$; acquisition mode is full scanning mode

Calculation of the concentration of volatile aroma components: n-amyl acetate is selected as the internal standard substance for quantitative analysis, and the internal standard solution is prepared according to *GB/T*10345 − 2007. The concentration and peak area of the internal standard substance are known, and the quantitative analysis is carried out according to the comparison of the peak area of target substance and the internal standard substance. The concentration is expressed as follows:

$$c = \frac{(A_1/A_2) \times m}{V}, \tag{1}$$

where c is the concentration of the aroma substance whose unit is mg/μL. A_1 and A_2 are the peak area of the aroma substance which require quantitative analysis and internal standard substance, respectively. m is the mass of the internal standard substance whose unit is mg. V is the volume of the Baijiu sample, and its unit is μL.

2.3. Infrared Spectrometric Measurement. Before spectral acquisition, all samples are stored in the laboratory at 4°C. Samples are scanned by using the Nicolet-6700 FT-NIR spectrometer (Thermo Fisher Scientific, USA) with the single-point attenuated total reflectance attenuation accessory under the room temperature 25 ± 0.5°C, and deionized water is utilized as the reference. The sample cuvette is cleaned more than three times by test samples and dried up before every measurement to refrain from pollution. Instrument parameters are provided as follows: spectral resolution is 4 cm^{-1}; measuring range is 4000 ~ 400 cm^{-1}; and successive scans times are 32. The spectra of each sample are corrected in triplicate, and the average value is regarded as the final spectral data.

Table 1: Distribution of Chinese spirit samples.

Type	Num. of samples	Training set	Test set
Raw spirit	20	15	5
1-year aged	20	15	5
3-year aged	20	15	5
5-year aged	20	15	5
Total	80	60	20

2.4. Spectral Data Pretreatment. In this paper, several spectral data pretreatment methods are employed, which are spectral smoothing, multivariate baseline correction, and first and second derivative, respectively. Spectral smoothing can reduce signal interference from high-frequency noise and improve the appearance of the spectrum. Since the baseline obtained in the spectrum may be tilted, drifted, or curved, baseline calibration is conducive to find desirable peaks, which is more profitable in spectral comparison or quantitative analysis. Multivariate baseline correction is a polynomial interpolation calculation for a specified baseline point, which is suitable for severely curved baselines. Furthermore, due to the coupling of different chemical groups in the Baijiu samples, the infrared absorption spectrum lines coincide. It is known that differential processing is proposed against the overlap of spectral lines. Consequently, the first derivative and the second derivative are commonly utilized. They can enhance the subtle spectral features. The first derivative is the rate of change of the whole spectrum, and the second is the change in the spectral rate change.

2.5. Principal Component Analysis. PCA is a multivariate statistical analysis method. The main principle is that the high-dimensional feature data are mapped to the low-dimensional space through orthogonal transformation. The linear independent variables in the low-dimensional space can contain the features of the original data, and the main components are defined. In general, the larger the signal data variance is, the greater is the amount of information contained in the signal. Because contained information mainly depends on the carrying characteristics of data variance, the cumulative variance contribution rate is employed to measure the amount of data information. The detailed steps are listed as follows:

Step 1: standardization of raw data: if there are m features and N samples in the original data, they can be expressed by the matrix of dimensions, that is,

$$\mathbf{X}_{N\times m} = \begin{bmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{Nm} \end{bmatrix}. \tag{2}$$

Step 2: the original data are normalized to generate the standard matrix $\mathbf{X}^*$ (the values of all elements are within 0 and 1), that is,

$$x_{ij}^* = \frac{x_{ij} - \overline{x}_j}{s_j}, \tag{3}$$

where $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, m$. $\overline{x}_j$, s_j are the mean value and variance of variable index x_j, respectively.

Step 3: the correlation matrix $\mathbf{R}$ of the standard matrix $\mathbf{X}^*$ in step 1 can be calculated by

$$\mathbf{R} = \frac{\mathbf{X}^{*T}\mathbf{X}^*}{(N-1)}. \tag{4}$$

Meanwhile, the eigenvalues of matrix $\mathbf{R}$ from large to small are calculated as $\lambda_1 > \lambda_2 > \cdots > \lambda_m$, and the corresponding eigenvectors can be also obtained as $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p, \ldots, \mathbf{u}_m$.

Step 4: determining the number of principal components: firstly, the variance contribution rate is calculated according to formula (5); then, the cumulative variance contribution rate can be obtained by equation (6).

$$\eta_i = \frac{\lambda_i}{\sum_i^m \lambda_i} \times 100\%, \tag{5}$$

$$\eta_{\sum}(P) = \sum_i^P \eta_i. \tag{6}$$

According to the cumulative variance contribution rate, the number of principal components can be determined. In general, the cumulative variance contribution rate of the selected main component should be within 80% and 97%, which can contain most of information of the original data.

Step 5: according to the principal components in step 3, it can be concluded that the corresponding eigenvector matrix is $\mathbf{U}_{m\times p} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p]$. Finally, the features of n samples are compressed to p principal components, and the dimensionality of the data is reduced. The matrix after dimension reduction is

$$Z_{N\times p} = X^*_{N\times m} U_{m\times p}. \tag{7}$$

2.6. Discriminant Analysis. Discriminant Analysis (DA) is a multivariate statistical analysis method for classification [39–42]. The basic principle of this method is that Baijiu samples are classified based on distance function, where the most commonly utilized method is the Mahalanobis distance. The Mahalanobis distance is calculated as

$$d = \sqrt{(x-\mu)^T S^{-1}(x-\mu)}, \tag{8}$$

where d is the Mahalanobis distance and x is the score vector of the sample. μ is the mean score vector of the sample sets, and S is the score covariance matrix. T is the transpose of $(x-\mu)$. Discriminant analysis is applied to calculate the Mahalanobis distance between unknown samples' spectrum and a set of standard spectra with TQ Analysis software. Consequently, those unknown samples will be classified to a given class and the Mahalanobis distance displayed for each class. The closer the value is to 0, the better the matching result is.

2.7. Support Vector Machine. For the training samples (x_i, y_i), $x_i \in \mathfrak{R}^n$ is regarded as the input of the SVM model and $y_i \in \mathfrak{R}$ is the output, where $i \in (1, 2, \ldots, N)$ is the number of the training samples. Throughout the nonlinear mapping $\phi(\cdot)$, the input data x_i can be mapped to a high-dimensional feature space. By the high-dimensional spatial map, a linearly nonseparable problem can be transformed into a linear separable problem in high-dimensional space, which is shown in Figure 1.

Hence, in this feature space, the regression model is mathematically expressed as

$$y = f(x) = \omega\phi(x) + b, \tag{9}$$

where ω is a weight vector and b is bias.

According to the principle of structural risk minimization, equation (9) can be rewritten as an optimization problem with equality constraints:

$$\begin{aligned} &\min_{\omega,b,\zeta} \frac{1}{2}\omega^T\omega + \frac{c}{2}\sum_{i=1}^N \zeta_i^2 \\ &\text{s.t.} \quad y_i = \omega^T\phi(x) + b + \zeta_i, \end{aligned} \tag{10}$$

where c is the regularization parameter. ζ_i is the relaxation variable.

The aforementioned problem (10) is a typical convex quadratic planning problem which can be solved by introducing Lagrange function. It can be expressed in detail as

$$L(\omega, b, \zeta, \alpha) = \frac{1}{2}\omega^T\omega + \frac{c}{2}\sum_{i=1}^N \zeta_i^2 - \sum_{i=1}^l \alpha_i\left[\omega^T\phi(x) + b + \zeta_i - y_i\right], \tag{11}$$

where $\alpha_i\,(i = 1, 2, \ldots, N)$ represents the Lagrange multiplier.

According to the optimal condition of Karush–Kuhn–Tucher (KKT), it can be concluded that

$$\begin{bmatrix} 0 & S^T \\ S & \dfrac{K+I}{c} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}, \tag{12}$$

where $S = [1, 1, \ldots, 1]^T$, $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]$, and $Y = [y_1, y_2, \ldots, y_N]$. It is worth noting that $K = \phi^T(x_k)\phi(x_N)$ is a kernel function satisfying the Mercer condition. In this paper, we adopt radial basis function as the kernel function of SVM. It is expressed in detail as

$$K(x, x_i) = \exp\left(\frac{-\left\|x - x_i\right\|^T}{2g^2}\right), \tag{13}$$

where g^2 represents the nuclear width.

The SVM classification model can be obtained by solving the linear equation (12). Also, the model is presented as

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^N \alpha_i K(x, x_i) + b\right). \tag{14}$$

From formula (14), we can conclude that the structure of SVM is similar to that of neural network. The output is a linear combination of intermediate nodes, and each

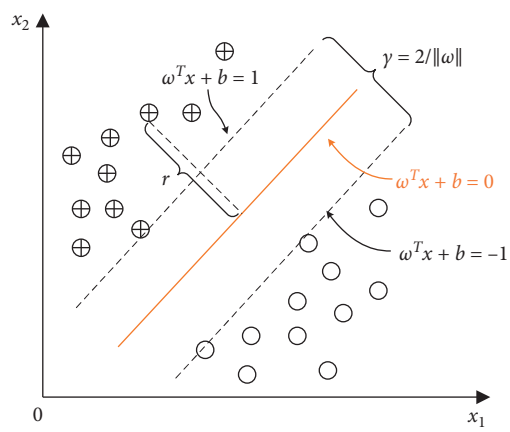

Figure 1: Architecture of SVM.

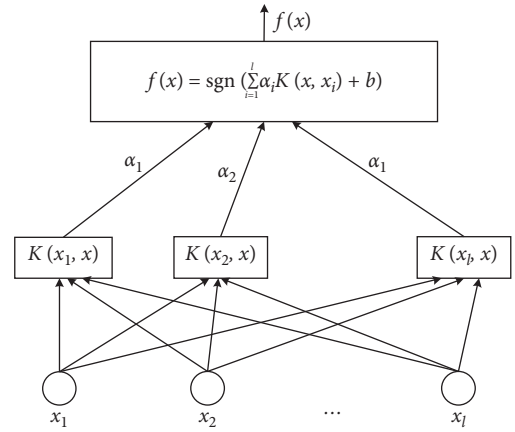

Figure 2: Support vector and interval.

intermediate node corresponds to a support vector. The schematic diagram of SVM is shown in Figure 2.

2.8. Statistical Analysis. Statistical treatment, including calculation of mean, relative standard deviation, and standard error, is performed with the STATISTICA 6.0 software (Stat Soft Inc., USA). Principal component analysis (PCA) and Discriminant Analysis (DA) are employed to evaluate the possible grouping of the Chinese Baijiu, by using the TQ Analysis Software, version 8.0, Thermo Fisher Scientific (USA). The modeling of Support Vector Machine (SVM) is completed by Matrix Laboratory (MATLAB), which can be utilized for qualitative modeling analysis, numerical calculation, and 3D drawing.

3. Statistical Analysis

3.1. Changes of Volatile Flavor Compounds during Spirit Storage. Acetic acid is one of the chief acids in Chinese Baijiu, and esters exist in the form of ethyl ester mostly. The component contents of ethyl caproate and ethyl lactate, which are related to the quality closely, are at a high level. They are the main aroma components of Luzhou-flavor Baijiu, which is consistent with the references. Changing regularities of the organic compounds are beneficial to explore the aging mechanism of Chinese Baijiu. It can be observed from Figure 3 that the major contents exhibit an increasing trend, a sharp growth tendency in the early stage and a mild growth in the later stage with the extension of storage time. Accordingly, we can infer that the physical and chemical reaction rate in Baijiu decreases and tends to be stable. Not only the content but also the types of substances have changed. Some new substances appeared such as propionic acid, valeric acid, hexyl hexanoate, ethyl decanoate, and so on. The reasons for their formation are the oxidation of alcohols, esterification of acids and corresponding alcohols, and hydrolysis of esters, which make all kinds of trace components to be in a dynamic equilibrium. The formation of new substances makes the Baijiu body become more abundant, which is indispensable in stabilizing and improving quality. To summarize, compared with base Baijiu samples, the aged Chinese Baijiu is more affluent in ingredients' content and variety. The change of the ratio of internal components and new substances makes the Baijiu body become more harmonious, which endows mellow taste and strong fragrances.

3.2. Original Spectral Analysis and Spectral Pretreatment. From Figure 4, it can be observed that the spectra of the four groups' samples are highly overlapped regardless of aging duration, which cannot be distinguished by naked eyes. Although there are hundreds of substances in the Chinese Baijiu, the MIR band consists of the base frequency and the fundamental absorption region of hydrogen-containing groups, which results in no significant difference on the whole except in the range of 2300 – 2400 cm^{-1}. Then, the wave band of 2300 – 2400 cm^{-1} is locally magnified and displayed in the medium-sized picture at the top right of Figure 4. The difference is visible after amplification, but the samples cannot be completely distinguished through original spectral analysis alone. The spectral data pretreatment, composed of spectral smoothing, multivariate baseline correction, and first and second derivative processing, are subsequently carried out to evaluate the classification of samples.

Compared with original spectra, the subtle differences can be significantly enhanced and amplified through derivative processing. Figures 5 and 6 are the results of first-order and second-order derivative spectral processing, respectively. Different from spectral smoothing and multivariate baseline correction, it makes the difference become more remarkable. The spectral characteristics of original spectrum in two bands of 2300 – 2400 cm^{-1} and 1400 – 1600 cm^{-1} are enhanced, and the absorption band at 1740 cm^{-1} is potentially related with esters. In addition, the absorption band at 1580 cm^{-1} might be related with Lactate [43–45]. However, it is difficult to distinguish them barely from the intensity, position, and shape of peak. Besides, the spectrum of Chinese Baijiu samples overlaps and interlaces, which makes the work become more challenging.

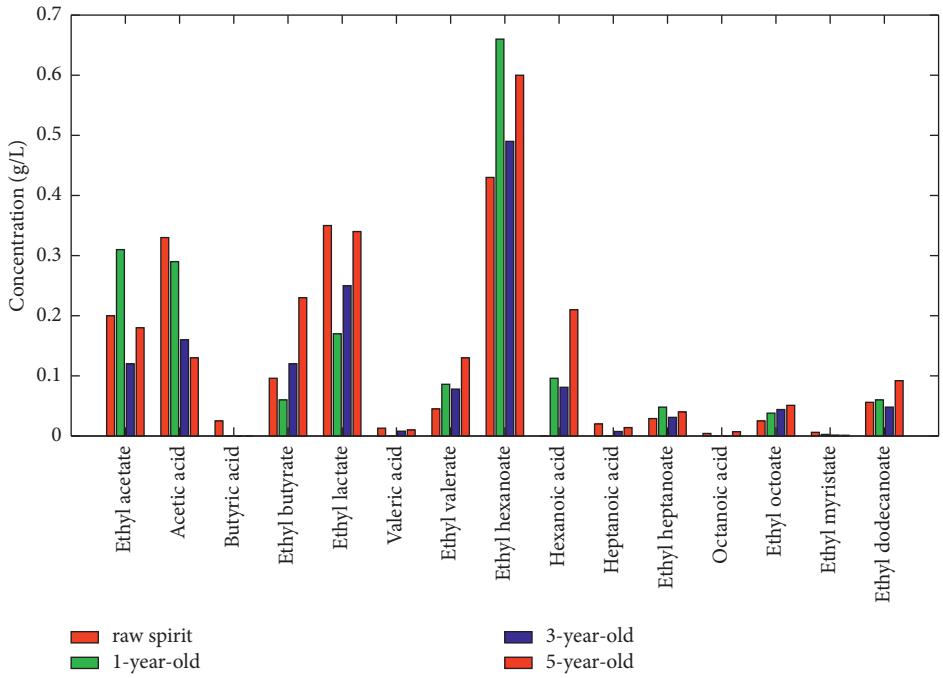

Figure 3: Instrumental analysis results of main volatile aroma components in the Chinese Baijiu from different ages (g/L). Independent colors indicate that there are significant differences among the four groups ($p < 0.05$).

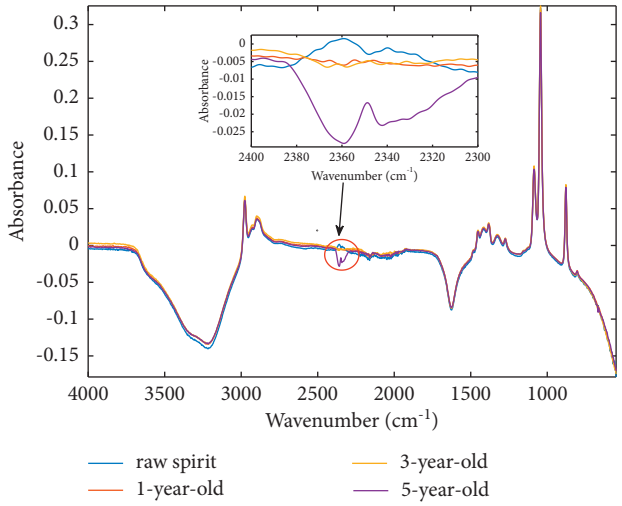

Figure 4: Original infrared spectrum of spirit samples in different aging times.

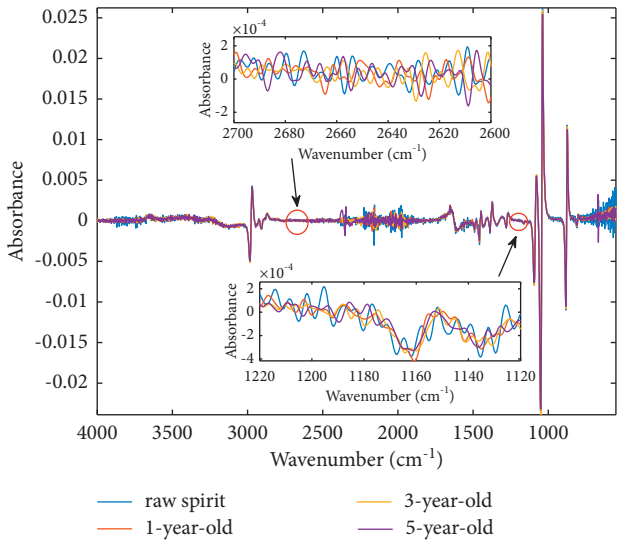

Figure 5: Infrared first-order differential spectroscopy of different age Chinese Baijiu samples.

3.3. PCA Analysis. The spectrum of wine age identification samples is collected on the whole band. The results of PCA are shown in Figures 7 and 8. From Figure 7, it can be seen that the later the component, the smaller the contribution rate of variance. The cumulative contribution rate of the first two principal components is as high as 99.8%, which is very close to 100%. PC1 and PC2 can represent the most of information of the infrared spectrum. From another perspective, it is impactful and feasible to utilize the means of PCA for dimension reduction.

Figure 8 shows the two-dimensional score figure of PC1 and PC2 derived from the original spectrum separately. It can be observed that boundaries between raw and aged Chinese Baijiu samples are very clear. The red marking part at the bottom right of the figure is samples of raw spirit, which are completely distinguished from aged samples. PCA technology exhibits the original spectrum of samples, that is, the characteristic information of the sample itself. The properties of them are quite different, and samples with different aging times (1, 3, and 5 years) are not completely discriminated, which indicates that their chemical attributes are not very alike. Five-year-old samples have an obvious clustering trend, but some of them overlap with 1-year-old

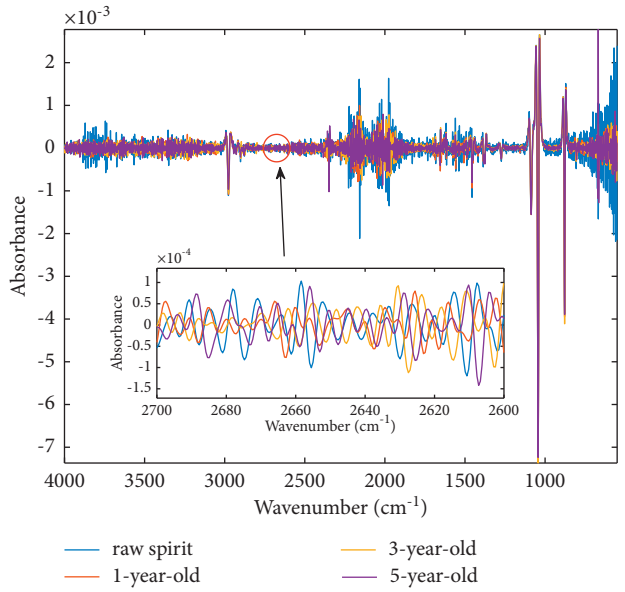

Figure 6: Infrared second-order differential spectroscopy of different age Chinese Baijiu samples.

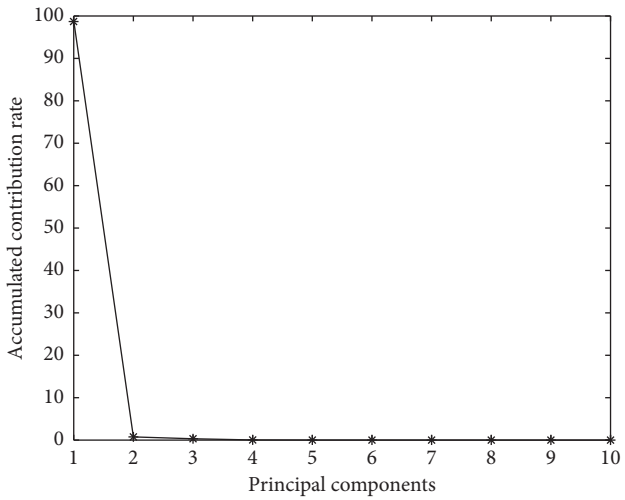

Figure 7: Cumulative contribution rate of the top 10 principal components in Baijiu samples.

samples. The black marking part of 5-year-old samples is inclined to cluster evidently, yet some overlap with 1-year-old samples. In the meantime, 1-year-old and 3-year-old samples are messy and hard to distinguish. That reason is the close nature of them. As for the whole, it is uncomplicated to distinguish whether the Baijiu is raw or mellow because of the great difference in its own properties. There are more or less overlapping phenomena in the aged Baijiu samples, especially in the 1-year-old and 3-year-old samples. They are approximate in terms of storage time, trace components, and spectral characteristics, which make the two become most likely confused.

3.4. DA Classification Results. PCA can merely achieve the distinction between raw and aged spirit samples. It is unrealistic to gain the complete classification of four kinds of substances. Therefore, the methods of discriminant analysis, different spectral bands, and spectral pretreatment ways are employed to establish the identification model, so as to avert the possible misjudgment caused by overlapping.

It is essential to select appropriate wavenumber for mitigating disturbance, improving prediction accuracy, and simplifying the model. According to the position of several main absorption peaks, the full spectrum ($594 - 3930\,cm^{-1}$) can be divided into three parts: $594 - 1042\,cm^{-1}$, $1042 - 1450\,cm^{-1}$, and $1450 - 3930\,cm^{-1}$, respectively. Table 2 is the classification results of Baijiu samples by DA in different spectral bands.

Table 2 shows model results from four distinct modeling bands: $594 - 3930\,cm^{-1}$, $594 - 1042\,cm^{-1}$, $1042 - 1450\,cm^{-1}$, and $1450 - 3930\,cm^{-1}$. The bands of $594 - 1042\,cm^{-1}$ have the worst results. Eleven Baijiu samples are misjudged where 7

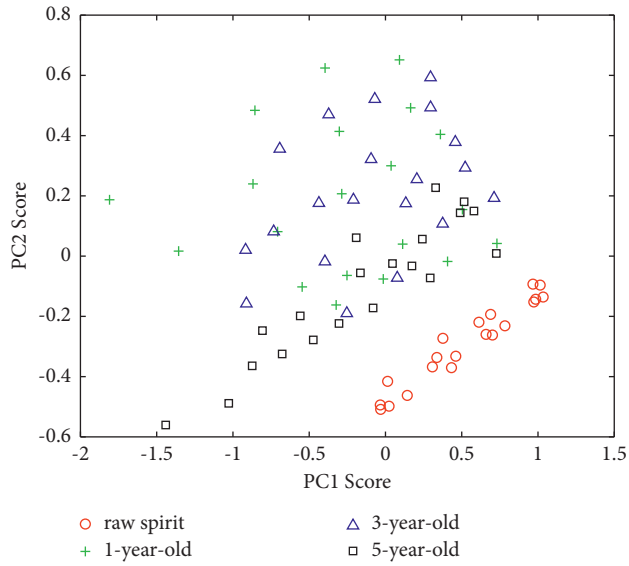

Figure 8: Principal component analysis score chart of different age Chinese Baijiu samples.

Table 2: Classification results of spirit samples by discriminant analysis in different spectral bands.

Waveband (cm^{-1})	Training set		Test set	
	Accuracy (percentage %)	Samples of misclassified	Accuracy (percentage %)	Samples of misclassified
594–3930	93.33	4	95.00	1
594–1042	83.33	7	80.00	4
1042–1450	91.67	5	80.00	4
1450–3930	91.67	5	85.00	3

samples are in the training set and 4 samples are from the test set. Poor discrimination results are acquired from two bands of $1042 - 1450\,cm^{-1}$ and $1450 - 3930\,cm^{-1}$, where 9 samples and 8 samples are misjudged separately. Furthermore, it is the least number of misjudgments in the full wavebands of $594 - 3930\,cm^{-1}$: 4 Baijiu samples are from the training set and 1 sample is in the test set. The accuracy of discrimination in the training set is 93.33% and 95.00% in the test set. The waveband of $594 - 3930\,cm^{-1}$ has achieved optimum results, which indicates that the spectrum, in the range, contains the key classification and identification information of Baijiu samples. According to the abovementioned analysis results, full band range is the final choice to modeling, and different spectral pretreatment methods are applied to complete the screening of them.

Table 3 shows model results from distinct spectral pretreatment methods: 5-point smoothing, 15-point smoothing, multivariate baseline correction, and first derivative and second derivative processing. It can be observed that the first-order and second-order differential processing have 15 and 11 misjudgment samples in amount, respectively, with the poor results. The prediction accuracy of first-order differential in the test set is merely 75.00% with great uncertainty. Yet, the results of smoothing and multivariate baseline correction are much better. Five samples are misjudged in total, which is consistent with the original spectral analysis results. In other words, smoothing and multivariate baseline correction processing have no essential changes on the treatment results. Comparing the results of differential processing and original spectral modeling, the number of misclassified samples in calibration sets decreases from 15 to 5. The accuracy of discrimination increases from 83.38% to 93.33% in the training set and increases from 75.00% to 95.00% in the test set. The decrease in the number of misclassification and improvement of the accuracy of discrimination are in the ideal direction. The qualitative identification model is established on the whole band combining with original spectrum finally.

Figures 9 and 10 are two-dimensional and three-dimensional Mahalanobis distance graphs of different liquor age samples based on the DA method. It can be seen that the raw Baijiu samples can be distinguished from mellow Baijiu samples evidently. It is more obvious in the three-dimensional Mahalanobis distance graph. From Figure 9, it can be observed that the raw Baijiu samples are located in the upper left of the graph far from aged spirit samples and the 5-year-old samples are at the bottom. The 3-year-old and 1-year-old samples are above them, where there exists an obvious clustering trend. The 3-year-old samples are in the left half, and the right half is 5-year-old spirit samples. On the whole, four miscalculations in the training set are that 1-year-old samples are miscalculated as 3-year-old samples. In the previous analysis of PCA, the characteristics of the 3-year-old

Table 3: Classification results of spirit samples by discriminant analysis in different spectral pretreatment methods.

Preprocessing methods	Training set		Test set	
	Accuracy (percentage %)	Samples of misclassified	Accuracy (percentage %)	Samples of misclassified
Raw spectra	93.33	4	95.00	1
5-point smoothing	93.33	4	95.00	1
15-point smoothing	93.33	4	95.00	1
Multivariate baseline correction	90.00	6	90.00	2
First derivative	83.38	10	75.00	5
Second derivative	86.66	8	85.00	3

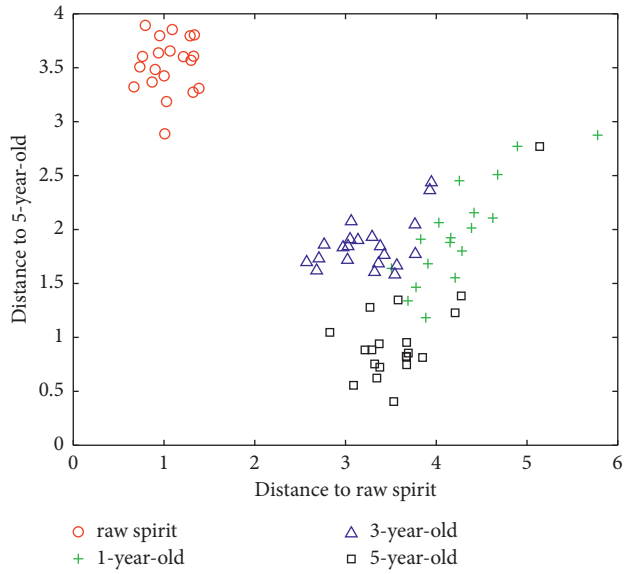

Figure 9: Two-dimensional Mahalanobis distance map of different liquor age samples based on the DA method (N = 80, performance index: 0.983).

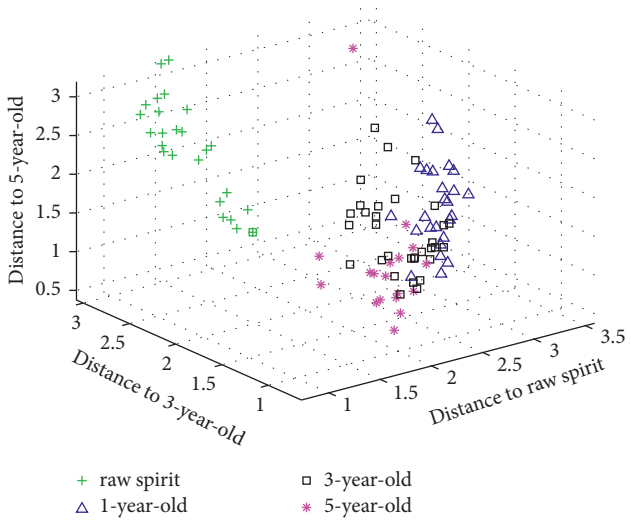

Figure 10: Three-dimensional Mahalanobis distance map of different liquor age samples based on the DA method (N = 80, performance index: 0.983).

samples are similar to those of 1-year-old samples due to the adjacent aging time, chemical properties, and spectral characteristics, which can be the explanation for the miscalculation. It is not difficult to distinguish the other samples because of the great difference in nature. Consequently, classification accuracies in the training and test sets

are 93.33% and 95.00%, respectively, by the DA method for the age classification of Chinese Baijiu.

4. SVM Classification Results

4.1. Parameter Optimization Based on Grid Search and Cross Validation. According to the principle of SVM, the regularization parameter c and kernel width parameter g^2 play an important role in the model. Consequently, before utilizing SVM to construct the Chinese Baijiu classification model, the regularization parameter c and kernel width parameter g^2 should be determined. In this paper, grid search (GS) and cross validation (CV) are employed to optimize the two parameters of SVM.

Grid search: the grid search method is an exhaustive method. This method takes several divisions in each dimension of the parameter space, and it traverses all grid intersections in the input space to obtain the optimal solution. The advantage of the grid search method is that it can ensure that the search solution is the global optimal solution in the delimited grid. Simultaneously, the significant errors can also be avoided. The details of the method are represented as follows: Firstly, to the best of our knowledge, the ranges of c and g^2 are set as $[-10, 10]$ to form a larger 2-dimensional plane. Then, based on this plane, the intervals of c and g^2 are divided into M points and N points at equal intervals to form an $M \times N$ grid plane. The intersection of the grid planes is a possible combination of parameters. Finally, for each parameter combination, the estimation error is calculated and the combination with the minimum error is the optimal parameter.

Cross validation: in this paper, the capacity of the samples is limited relatively. In order to make full use of all the sample dataset for training and test, the cross-validation method is employed by minimizing the mean square error (MSE), which is expressed as

$$\mathrm{MSE} = \frac{1}{l}\sum_{i=1}^{l}\left(y_i - \widehat{y}_i\right)^2, \tag{15}$$

where y_i and $\widehat{y}_i$ are the actual value and estimation value, respectively.

As a matter of fact, it is worth noting that the SVM classification performance of parameter combination is affected by the training data. For the same group (C, g^2), when the training data change, the corresponding SVM performance also changes. In particular, considering small sample training, the parameter optimization is greatly affected by the randomness of the sample, which is not conducive to the generalization and promotion of the model. Based on the abovementioned discussion, the k-fold cross validation method is adopted to comprehensively evaluate the performance of each group (C, g^2).

4.2. PCA-GS-CV-SVM Classification Model. The qualitative identification analysis model of Chinese Baijiu samples is established based on the SVM algorithm in libsvm toolbox of MATLAB. The specific steps are as follows:

Step 1: PCA of the infrared spectra of all samples is carried out over the full spectra range.

Step 2: the data after the PCA are divided into the training dataset and test dataset. Establishing the correspondence between sample categories and labels simultaneously, the corresponding relationships are listed as follows: raw spirit (1); 1 year old (2); 3 years old (3); and 5 years old (4).

Step 3: the input and output data from the training set, together with input data from the test set, are normalized. The normalization formula is as follows:

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \tag{16}$$

where y_i is the normalized data, x_i is the original data, and $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the original data, respectively.

Step 4:eEstablishing and training the qualitative model of SVM: The radial basis function is used in this paper to obtain better qualitative accuracy, and the cross-validation method is used to find the optimal SVM model parameters, including the penalty factor c and the variance g in the radial basis function.

Step 5: the input data from the test set are input to the trained SVM qualitative model to detect the performance of the established model.

To explain the principle and scheme of the PCA-GS-CV-SVM classification model, the entire frame is given by Figure 11.

4.3. Results Analysis. As shown in Figure 7, the accumulated contribution rate of the first three principal components is 99.8%, close to 100%. The contribution rates of the latter components are small. Most of the spectral information is represented by the first three principal components. Therefore, it can be considered that PCA is reliable for reducing dimensionality of the Chinese Baijiu identification samples. When the penalty factor c is 0.5000 and the variance in the radial basis function g is 0.2176, the qualitative model of SVM combined with PCA is established. The identification results are shown in Figures 12 and 13. In addition, the classification results of Baijiu samples by different models are given in Table 4. It can be seen that a total of 100% classification accuracy is obtained in the training set

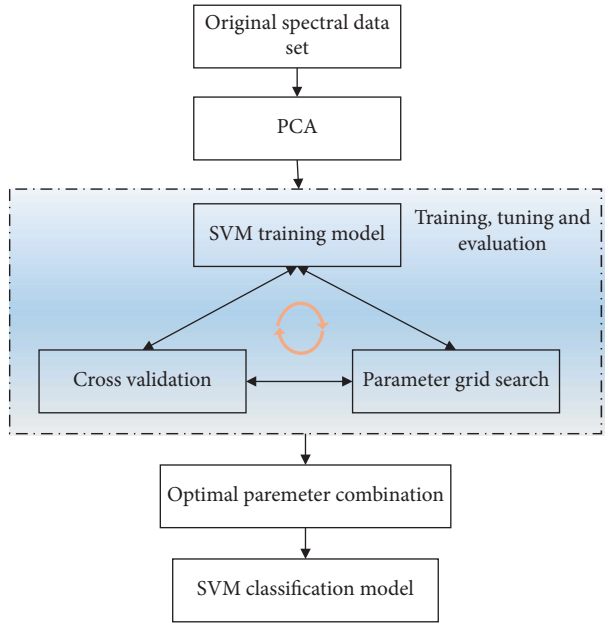

Figure 11: PCA-GS-CV-SVM classification model.

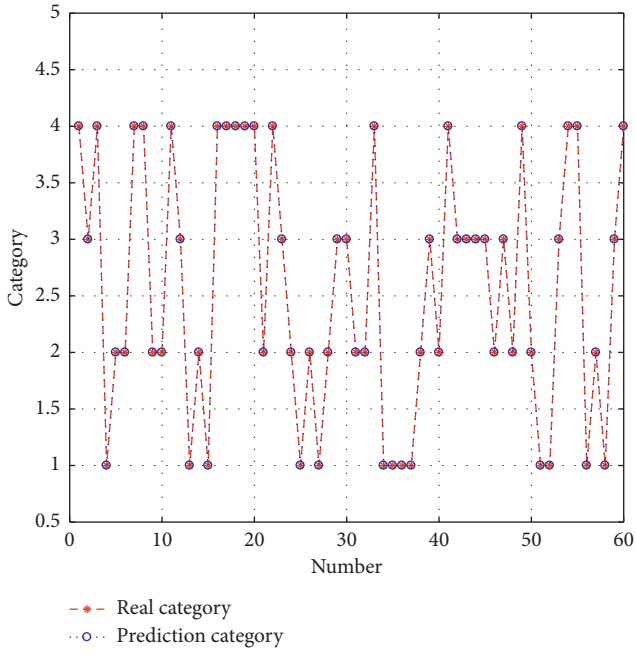

Figure 12: SVM qualitative results comparison of the training set (N = 60, c = 0.5000, and g = 0.2176).

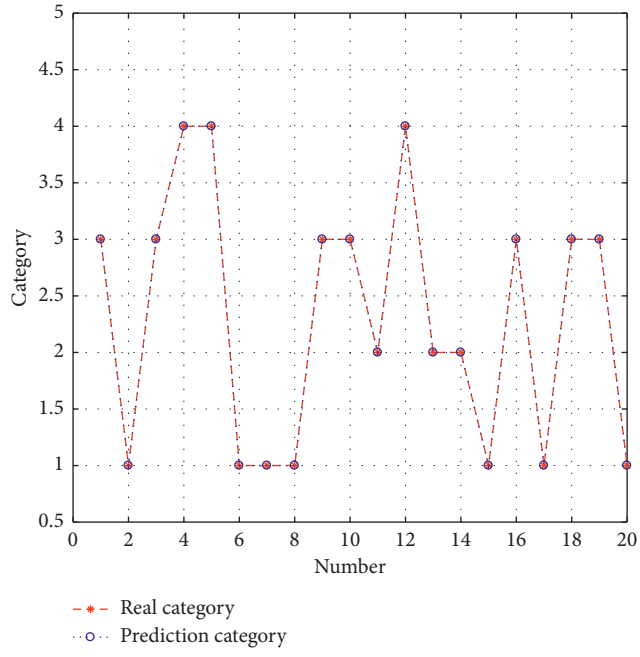

Figure 13: SVM qualitative results comparison of the test set (N = 20, c = 0.5000, and g = 0.2176).

Table 4: Classification results of spirit samples by different models.

Preprocessing methods	Calibration set		Prediction set	
	Accuracy (percentage %)	Samples of misclassified	Accuracy (percentage %)	Samples of misclassified
PCA-BP	93.33	4	95.00	1
PCA-RBF	90.00	6	90.00	2
PCA-SVM	96.67	2	95.00	1
PCA-GS-CV-SVM	100	0	100	0

and test set. The classification of the established model is completely consistent with the actual ascription, which shows that the SVM model could distinguish the different age groups excellently.

5. Conclusions

In this paper, we propose a liquor age discrimination method of Chinese Baijiu based on midinfrared spectroscopy and chemometrics. Meanwhile, the identifying results are demonstrated based on different modeling methods, spectral preprocessing, and band selection. Five-point, 15-point spectral smoothing, and multivariate baseline correction have little effect on the analysis results, but the derivative processing is the worst. As far as the modeling method is concerned, PCA can merely achieve the distinction between raw and aged Chinese Baijiu samples. It is unrealistic to obtain the complete classification of four kinds of substances. The DA method mistakenly judges 1-year-old and 5-year-old samples as 3-year-old with 93.33% classification accuracy in the training set and 95% in the test set. A total of 100% classification accuracy is obtained in the training set and test set by employing PCA-GS-CV-SVM algorithm. This method can obtain ideal experimental results and can be applied for the rapid and nondestructive detection of Chinese Baijiu. However, the current work focuses on the liquor age classification of Luzhou-flavor Baijiu, one of the classic flavors of Chinese Baijiu, and the number of samples is limited. In further research, added samples should be collected from different flavors, regions, and grades to establish a more complete calibration model.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Postgraduate Research and Practice Innovation Program of Jiangsu Province (Nos. KYCX20-0572 and KYCX20-0207) and Interdisciplinary Innovation Foundation for Graduates, NUAA (No. KXKCX JJ202008).

References

[1] Y. Li, S. Fan, A. Li et al., "Vintage analysis of Chinese baijiu by GC and 1H NMR combined with multivariable analysis," *Food Chemistry*, vol. 360, Article ID 129937, 2021.

[2] A. Bai, S. Liu, A. Chen et al., "Residue changes and processing factors of eighteen field-applied pesticides during the production of Chinese baijiu from rice," *Food Chemistry*, vol. 359, Article ID 129983, 2021.

[3] Q. Zheng, Z. Wang, A. Xiong et al., "Elucidating oxidation-based flavour formation mechanism in the aging process of Chinese distilled spirits by electrochemistry and UPLC-Q-Orbitrap-MS/MS," *Food Chemistry*, vol. 355, Article ID 129596, 2021.

[4] T. Shen, Q. Wu, and Y. Xu, "Biodegradation of cyanide with saccharomyces cerevisiae in baijiu fermentation," *Food Control*, vol. 127, Article ID 108107, 2021.

[5] Y. Zhang, J. Gu, C. Ma et al., "Flavor classification and year prediction of Chinese baijiu by time-resolved fluorescence," *Applied Optics*, vol. 60, no. 19, pp. 5480–5487, 2021.

[6] X. He, H. Yangming, E. Górska-Horczyczak, A. Wierzbicka, and H. H. Jeleń, "Rapid analysis of baijiu volatile compounds fingerprint for their aroma and regional origin authenticity assessment," *Food Chemistry*, vol. 337, Article ID 128002, 2021.

[7] W. Jia, Z. Fan, A. Du et al., "Recent advances in Baijiu analysis by chromatography based technology-a review," *Food Chemistry*, vol. 324, Article ID 126899, 2020.

[8] X. Song, S. Jing, L. Zhu et al., "Untargeted and targeted metabolomics strategy for the classification of strong aroma-type baijiu (liquor) according to geographical origin using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry," *Food Chemistry*, vol. 314, Article ID 126098, 2020.

[9] H. Li, D. Qin, Z. Wu et al., "Characterization of key aroma compounds in Chinese guojing sesame-flavor baijiu by means of molecular sensory science," *Food Chemistry*, vol. 284, pp. 100–107, 2019.

[10] Q. Peng, X. Xu, W. Xing et al., "Ageing status characterization of Chinese spirit using scent characteristics combined with chemometric analysis," *Innovative Food Science and Emerging Technologies*, vol. 44, pp. 212–216, 2017.

[11] J. Zheng, R. Liang, J. Huang et al., "Volatile compounds of raw spirits from different distilling stages of luzhou-flavor spirit," *Food Science and Technology Research*, vol. 20, no. 2, pp. 283–293, 2014.

[12] D. M. Musingarabwi, H. H. Nieuwoudt, P. R. Young, H. A. Eyéghè-Bickong, and M. A. Vivier, "A rapid qualitative and quantitative evaluation of grape berries at various stages of development using fourier-transform infrared spectroscopy and multivariate data analysis," *Food Chemistry*, vol. 190, pp. 253–262, 2016.

[13] L. Liu, I. Loira, A. Morata et al., "Shortening the ageing on lees process in wines by using ultrasound and microwave

treatments both combined with stirring and abrasion techniques," *European Food Research and Technology*, vol. 242, no. 4, pp. 559–569, 2016.

[14] Y. Tao, J. F. García, and D.-W. Sun, "Advances in wine aging technologies for enhancing wine quality and accelerating wine aging process," *Critical Reviews in Food Science and Nutrition*, vol. 54, no. 6, pp. 817–835, 2014.

[15] F. M. Agazzi, J. Nelson, C. K. Tanabe, C. Doyle, R. B. Boulton, and F. Buscema, "Aging of malbec wines from Mendoza and California: evolution of phenolic and elemental composition," *Food Chemistry*, vol. 269, pp. 103–110, 2018.

[16] M. Razeghi and B.-M. Nguyen, "Advances in mid-infrared detection and imaging: a key issues review," *Reports on Progress in Physics*, vol. 77, no. 8, Article ID 082401, 2014.

[17] J. Peng, "Developments of mid-infrared optical parametric oscillators for spectroscopic sensing: a review," *Optical Engineering*, vol. 53, no. 6, Article ID 061 613, 2014.

[18] M. Ye, Z. Gao, Z. Li, Y. Yuan, and T. Yue, "Rapid detection of volatile compounds in apple wines using ft-nir spectroscopy," *Food Chemistry*, vol. 190, pp. 701–708, 2016.

[19] N. Zhang, X. Liu, X. Jin et al., "Determination of total iron-reactive phenolics, anthocyanins and tannins in wine grapes of skins and seeds based on near-infrared hyperspectral imaging," *Food Chemistry*, vol. 237, pp. 811–817, 2017.

[20] J. L. Aleixandre-Tudo, H. Nieuwoudt, J. L. Aleixandre, and W. du Toit, "Chemometric compositional analysis of phenolic compounds in fermenting samples and wines using different infrared spectroscopy techniques," *Talanta*, vol. 176, pp. 526–536, 2018.

[21] C. Condurso, F. Cincotta, G. Tripodi, and A. Verzera, "Characterization and ageing monitoring of marsala dessert wines by a rapid ftir-atr method coupled with multivariate analysis," *European Food Research and Technology*, vol. 244, no. 6, pp. 1073–1081, 2018.

[22] Y. Deng, M. Chen, G. Chen et al., "Visible-ultraviolet upconversion carbon quantum dots for enhancement of the photocatalytic activity of titanium dioxide," *ACS Omega*, vol. 6, no. 6, pp. 4247–4254, 2021.

[23] J. Huang and K. Pu, "Activatable molecular probes for second near-infrared fluorescence, chemiluminescence, and photoacoustic imaging," *Angewandte Chemie-International Edition*, vol. 59, no. 29, pp. 11717–11731, 2020.

[24] Z. Genisheva, C. Quintelas, D. P. Mesquita, E. C. Ferreira, J. M. Oliveira, and A. L. Amaral, "New pls analysis approach to wine volatile compounds characterization by near infrared spectroscopy (nir)," *Food Chemistry*, vol. 246, pp. 172–178, 2018.

[25] B. Peng, N. Ge, L. Cui, and H. Zhao, "Monitoring of alcohol strength and titratable acidity of apple wine during fermentation using near-infrared spectroscopy," *Lebensmittel-Wissenschaft und -Technologie- Food Science and Technology*, vol. 66, pp. 86–92, 2016.

[26] S. Chen, F. Zhang, J. Ning, X. Liu, Z. Zhang, and S. Yang, "Predicting the anthocyanin content of wine grapes by nir hyperspectral imaging," *Food Chemistry*, vol. 172, pp. 788–793, 2015.

[27] C. A. Teixeira Dos Santos, R. N. M. J. Páscoa, M. C. Sarraguça et al., "Merging vibrational spectroscopic data for wine classification according to the geographic origin," *Food Research International*, vol. 102, pp. 504–510, 2017.

[28] M. Basalekou, C. Pappas, Y. Kotseridis, P. A. Tarantilis, E. Kontaxakis, and S. Kallithraka, "Red wine age estimation by the alteration of its color parameters: fourier transform infrared spectroscopy as a tool to monitor wine maturation time," *Journal of Analytical Methods in Chemistry*, vol. 2017, Article ID 5767613, 9 pages, 2017.

[29] S.-Y. Li, B.-Q. Zhu, M. J. Reeves, and C.-Q. Duan, "Phenolic analysis and theoretic design for Chinese commercial wines authentication," *Journal of Food Science*, vol. 83, no. 1, pp. 30–38, 2018.

[30] F. Shen, Y. Ying, B. Li, Y. Zheng, and X. Liu, "Discrimination of blended Chinese rice wine ages based on near-infrared spectroscopy," *International Journal of Food Properties*, vol. 15, no. 6, pp. 1262–1275, 2012.

[31] F. Shen, D. Yang, Y. Ying, B. Li, Y. Zheng, and T. Jiang, "Discrimination between Shaoxing wines and other Chinese ricewines by near-infrared spectroscopy and chemometrics," *Food and Bioprocess Technology*, vol. 5, no. 2, pp. 786–795, 2012.

[32] N. Gerhardt, S. Schwolow, S. Rohn et al., "Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: comparison of different processing approaches by LDA, KNN, and SVM," *Food Chemistry*, vol. 278, pp. 720–728, 2019.

[33] N. Zhu, K. Wang, S.-L. Zhang, B. Zhao, J.-N. Yang, and S.-W. Wang, "Application of artificial neural networks to predict multiple quality of dry-cured ham based on protein degradation," *Food Chemistry*, vol. 344, Article ID 128586, 2021.

[34] S. Abdullah, R. C. Pradhan, D. Pradhan, and S. Mishra, "Modeling and optimization of pectinase-assisted low-temperature extraction of cashew apple juice using artificial neural network coupled with genetic algorithm," *Food Chemistry*, vol. 339, Article ID 127862, 2021.

[35] J. Gao, L. Zhao, J. Li, L. Deng, J. Ni, and Z. Han, "Aflatoxin rapid detection based on hyperspectral with 1d-convolution neural network in the pixel level," *Food Chemistry*, vol. 360, Article ID 129968, 2021.

[36] H. S. Green, X. Li, M. de Pra et al., "A rapid method for the detection of extra virgin olive oil adulteration using UHPLC-CAD profiling of triacylglycerols and PCA," *Food Control*, vol. 107, Article ID 106773, 2020.

[37] O. Devos, G. Downey, and L. Duponchel, "Simultaneous data pre-processing and SVM classification model selection based on a parallel genetic algorithm applied to spectroscopic data of olive oils," *Food Chemistry*, vol. 148, pp. 124–130, 2014.

[38] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[39] Y. Zhou, Z. Chen, T. Liu, and J. Mao, "Extended canonical variates analysis for wine origin discrimination by using infrared spectroscopy," *Chemistry Letters*, vol. 45, no. 5, pp. 564–566, 2016.

[40] D.-Y. Kim, B.-K. Cho, S. H. Lee, K. Kwon, E. S. Park, and W.-H. Lee, "Application of fourier transform-mid infrared reflectance spectroscopy for monitoring Korean traditional rice wine makgeolli fermentation," *Sensors and Actuators B: Chemical*, vol. 230, pp. 753–760, 2016.

[41] J. Yu, H. Wang, J. Zhan, and W. Huang, "Review of recent UV-vis and infrared spectroscopy researches on wine detection and discrimination," *Applied Spectroscopy Reviews*, vol. 53, no. 1, pp. 65–86, 2018.

[42] J. Zhong, J. Chen, L. Yao, and T. Pan, "Discriminant analysis of liquor brands based on moving-window waveband screening using near-infrared spectroscopy," *American Journal of Analytical Chemistry*, vol. 9, no. 3, pp. 124–133, 2018.

[43] R. Banc, F. Loghin, D. Miere, F. Fetea, and C. Socaciu, "Romanian wines quality and authenticity using ft-mir

spectroscopy coupled with multivariate data analysis," *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, vol. 42, no. 2, pp. 556–564, 2014.

[44] D. Markechová, P. Májek, and J. Sádecká, "Fluorescence spectroscopy and multivariate methods for the determination of brandy adulteration with mixed wine spirit," *Food Chemistry*, vol. 159, pp. 193–199, 2014.

[45] G. Hernández, R. León, and A. Urtubia, "Detection of abnormal processes of wine fermentation by support vector machines," *Cluster Computing*, vol. 19, no. 3, pp. 1219–1225, 2016.