Hindawi
Journal of Food Quality
Volume 2024, Article ID 6217243, 10 pages
https://doi.org/10.1155/2024/6217243

Research Article

Application of Near-Infrared Spectroscopy to Rapidly Classify the Chinese Quince Fruits from Different Habitats

Songfeng Diao,[1,2] **Xiaoqian Tang**,[3] **Lin Huang**,[1] **Yanjie Li**,[2] **Xiongfei Fan**,[1] **and Wenhao Shao**[2]

[1]*Research Institute of Non-Timber Forestry, Chinese Academy of Forestry, Key Laboratory of Non-Timber Forest Germplasm Enhancement & Utilization of National Forestry and Grassland Administration, Zhengzhou 450003, Henan, China*
[2]*Institute of Subtropical Forestry Research of Chinese Academy of Forestry, Key Laboratory of Subtropical Tree Breeding and Cultivation of State Forestry and Grassland Administration, Key Laboratory of Tree Breeding of Zhejiang Province, Hangzhou 311400, Zhejiang, China*
[3]*Science and Technology Management Department, Chinese Academy of Forestry, Beijing 100091, China*

Correspondence should be addressed to Wenhao Shao; whshao8119@163.com

Received 7 March 2023; Revised 31 October 2023; Accepted 24 January 2024; Published 7 February 2024

Academic Editor: Timothy Omara

Copyright © 2024 Songfeng Diao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The ecological habitats of Chinese quince (*Chaenomeles speciosa* Nakai) fruits affect their phenotype. Currently, limited or no rapid method exists for classifying Chinese quince fruit from different ecosystems. This study developed a partial least squares discriminant analysis (PLS-DA) classification model to effectively and nondestructively classify 663 Chinese quince fruit samples from six environments in 2020. PLS-DA models and other variable selection approaches were used in this study. The near-infrared spectroscopy (NIRs) absorption spectra of raw Chinese quince fruit samples from six habitats showed a similar shape. The spectra of each environment showed little variance. The raw fruit spectra varied significantly among habitat categories after the first derivative preprocessing phase. The uninformative variable elimination (UVE) variable selection approach had greater calibration and validation set specificity of 0.93 and 0.98. This study found the best classification specificity using the UVE variable selection approach compared to other methods including the PLS-DA model without variable selection. The UVE approach improved Yunnan habitat categorization specificity from 86% to 88% when integrated with PLS-DA. Additionally, the validation set for quinces originating from Anhui, Chongqing, Hubei, Shandong, and Zhejiang achieved an ideal classification score of 100%. The findings of the study indicated that PLS-DA can serve as an alternative approach for classifying the habitats of Chinese quince fruits. When used in conjunction with other methods, this technique can assist researchers, scientists, and industry professionals in identifying the main factors responsible for significant variations in the habitats, composition, and quality of Chinese quince fruits.

1. Introduction

China is the natural habitat and cultivation center of Chinese quince (*Chaenomeles speciosa* Nakai), which has vast genetic resources and is mostly planted in the East, Central, and Southwest regions of China [1]. China is also the second-largest producer of quince in worldwide production, following Turkey. This fruit is a rich source of nutritious components, and it also possesses antioxidant and immune regulatory qualities. Sugar, amino acids, flavonoids, saponins, organic acids, and other useful components can be found in the fruit, which also possesses the ability to relax channels, activate collaterals, moisten the stomach, and perform a variety of other functions [2, 3]. It has also been used for thousands of years as one of the most essential substances in traditional Chinese medicine, which is typically appropriate to treat several diseases, including arthralgia, leg edema, and sunstroke [4]. To the present, Chinese quinces have continued to get a growing amount of attention for their potential to improve one's overall health. However, the quality of Chinese

quince fruit might change depending on the habitat where it is grown because of the varying climatic circumstances (such as the moisture and humidity levels of the soil and the temperature). Therefore, there is a growing demand for research to determine the quality of Chinese quince fruits grown in a variety of field conditions.

There are several reports that have been published in the past that discuss various methods, traditional and emerging, that have been used to determine the quality of fruits that have been produced in a range of different field conditions. Traditional methods, such as DNA analysis [5], amino acid composition [6], and gas chromatography (GC) analysis [7], were both time-consuming and expensive. Near-infrared spectroscopy (NIRs) is an emerging method that is both rapid and nondestructive [8]. It is used for qualitative and quantitative analysis of the chemical composition of fruits such as apples [9], bananas [10], peaches [11], kiwifruits [12], and pears [13]. This method is becoming increasingly popular as a solution to the limitations and challenges of traditional methods. NIRs has been used to authenticate the authenticity of freeze-dried açai pulp [14], trace apple habitat [15], determine soluble solid content in multihabitat apples [16], differentiate apple varieties, and investigate organic status [17]. Nevertheless, despite the number of research on fruit quality and habitat as discussed in the preceding lines, there is very little or no known research work related to the use of NIRs to determine Chinese quince habitat. In our earlier research [2], we analyzed and compared three distinct methods of discriminant analysis to determine the Chinese quince habitat.

Partial least squares discriminant analysis (PLS-DA) is one of the most widely used methods for classification in chemometrics [18, 19]. This method has also received widespread application in domains associated with the "omics," such as metabolomics, proteomics, and genomics, in addition to an array of other fields that generate huge amounts of data, such as spectroscopy [20–24]. The rising interest in PLS-DA, particularly in the field of metabolomics, may largely be attributed to the fact that it is included in the vast majority of widely used statistical software programs [22, 25–30]. These software packages include R, S-Plus, SAS, SPSS, and MATLAB. On the other hand, PLS-DA has recently been described by researchers as a powerful and reliable classification approach when paired with spectroscopy, which is utilized for discriminating between different qualities of fruit [31–33]. However, the PLS algorithm has a flaw in that it might provide inaccurate predictions due to the large number of irrelevant variables that it considers [34]. The methods used for selecting variables can choose a limited number of variables that are extremely significant and have an association with the characteristics of the class (for example, habitat) [35]. Variable selection may also increase classification performance by accurately selecting a subset of key predictors [36]. This can be done by using the results of the classification.

The utilization of NIRs has recently been employed to efficiently categorize Chinese quince fruits originating from distinct habitats [2]. The NIRs method provides a non-invasive and highly effective approach for analyzing the chemical composition of fruit samples [2, 8, 9, 14, 15, 31]. In a scientific investigation, scientists employed near-infrared reflectance spectroscopy in conjunction with multivariate analysis methodologies to categorize Chinese quince fruits according to their specific geographical origins [2]. The current investigation centered on Chinese provinces renowned for their diverse climate conditions and soil characteristics. The objective of the preceding investigation was to construct a model capable of effectively discriminating quince fruits originating from the aforementioned two geographical areas [2]. The investigation gathered NIRs spectra from a substantial quantity of quince fruit samples and employed multivariate analysis techniques, including principal component analysis (PCA) and linear discriminant analysis (LDA), to categorize the samples. The PCA was employed to effectively decrease the dimensionality of the spectral data. Subsequently, the LDA was utilized to construct a classification model using the reduced dataset. The findings of the research demonstrated that the NIRs methodology, in conjunction with multivariate analysis techniques, exhibited a high level of efficacy in accurately categorizing Chinese quince fruits originating from diverse habitats. Consequently, the classification accuracy exhibited a notable level, suggesting that NIRs possesses significant potential as a valuable instrument for swiftly and non-invasively categorizing fruit samples according to their geographical origin or natural habitat [2]. The utilization of NIRs in the categorization of Chinese quince fruits originating from diverse habitats showcases the promising capabilities of this method in ensuring fruit quality control, traceability, and authentication within the agricultural sector.

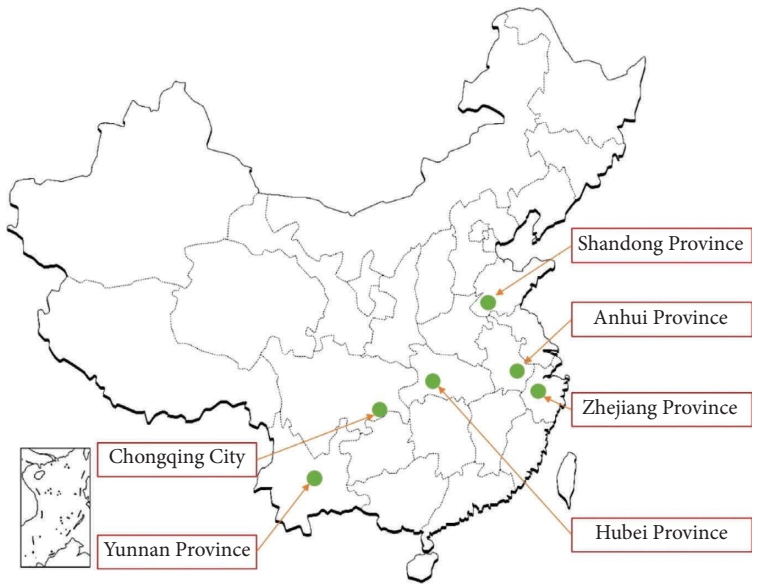
Therefore, the study aimed to develop PLS-DA models based on the NIRs of Chinese quince fruits to predict the habitats of Chinese quince and demonstrate how different variable selection methods influence the classification results of PLS-DA models rapidly and accurately.

2. Materials and Methods

2.1. Materials. During the harvest season in the year 2020, samples of Chinese quince fruit were collected from six different habitats (Figure 1), which together represent the majority of the Chinese quince fruit-producing regions. When the fruit's color changed to a yellowish green, which is also the customary time for harvesting quinces for medicinal purposes, three fresh quinces that were still intact were picked at random from each tree in each habitat. All of the samples were thereafter placed in a plastic bag, which was then labeled and then placed in a cooler box to maintain their freshness. The samples for the test consisted of a total of 663 fruits, which were collected from six main producing regions at a rate of three fruits per plant for a total of 221 distinct plants (Table 1 and Figure 2).

2.2. Methods

2.2.1. Spectra Acquisition. In this study, the data for the near-infrared reflectance spectra of individual fruits were



Habitats	Climatic Conditions	Geographical Conditions	
		Longitude	Latitude
Shandong	Warm-temperate monsoonal	114° 36′–112° 43′ E	34° 25′–38° 23′ N
Anhui	Warm-temperate	116° 09′–118° 10′ E	33° 18′–34° 38′ N
Zhejiang	Humid subtropical	118° 01′–123° 10′ E	27° 02′–31° 11′ N
Hubei	Humid subtropical	108° 21′–116° 07′ E	29° 05′–33° 20′ N
Chongqing	Warm and temperate	105° 11′–110° 11′ E	28° 10′–32° 13′ N
Yunnan	Plateau monsoon	97° 31′–106° 11′ E	21° 08′–29° 15′ N

Figure 1: Mapping presentation of sample collection sites of Chinese quince fruits from six different habitats.

Table 1: Different locations for the sample collection of Chinese quince fruit.

Habitats	Total (trees)	Total (fruit samples)
Shandong	22	66
Anhui	43	129
Zhejiang	33	99
Hubei	53	159
Chongqing	42	126
Yunnan	28	84
Total	221	663

collected at room temperature (25°C) using a hand-held near-infrared spectrometer (LF-2500, Spectral Evolution, USA) at an interval of 6 nm from 1000 nm to 2500 nm. A total of 32 times, on average, were used for scanning each spectrum. The manufacturer of the apparatus supplied the DARWin SP (version 1.2) software that was used to analyze the collected data.

Each individual fruit sample was subjected to the recording of all three spectra. The contact probe, which had a diameter of 20 mms, was positioned on the ventral surface of the Chinese quince fruit samples with the stem-calyx axis horizontal at a location chosen at random. The second measurement was carried out at a location that was roughly 120° rotated from the starting point. The third spectra were collected at an angle of roughly 240° rotated from the starting point. For each sample, an average of the three spectra was calculated.

2.2.2. Data Processing. The R software (version 3.1.2) was utilized for the processing of the data [37]. The NIRs spectra were averaged using the mean value of all of the fruits that were found on each tree. In the end, 221 different spectral samples were utilized. Following the conversion of the reflectance spectrum into the absorbance spectrum, multivariate analysis was performed. Both the standard normal variable and the first derivative were put through their tests as potential spectral preprocessing methods. The additive effect and noise present in the spectrum can be effectively eliminated through the utilization of two distinct preprocessing techniques, which differ from the conventional methods employed for processing NIRs spectra [2, 14].

The dataset was subsequently partitioned into two distinct subsets: a calibration set and a validation set [14]. Both of these subsets comprised samples that were chosen interactively using their Euclidean distances, aiming to achieve the highest attainable data coverage. Ultimately, a total of



Figure 2: Chinese quince fruits from six different habitats.

181 samples were employed for the calibration set, while the remaining 40 samples were allocated for the validation set [34].

PLS-DA classification models were utilized to differentiate between the various origins of Chinese quince fruits [17]. The PLS-DA method is a variant of the PLS regression (PLS-R) methodology. PLS-R is usually used to tackle regression-related problems and is most appropriate in situations in which the matrix of predictors contains more variables than data. PLS-DA is an appropriate approach for classification since it conducts a dimension reduction on the predictor variables and extracts the components that are significantly linked with the class factor [14, 16]. As a result, PLS-DA was employed to classify data.

In the PLS-DA model, the spectra of the six different habitat fruits were utilized for the X matrix, and six fabricated values were used for the Y matrix to represent each habitat. Shandong, Anhui, Zhejiang, Hubei, Chongqing, and Yunnan each had a dummy value between 0 and 5, and those values were given to their respective spectra. Root mean square error (RMSE) ranges of ±0.5 were set between each habitat. If an individual's RMSE fell within one of these ranges from any habitat, then the individual was considered to be classified in that habitat. The leave-one-out cross-validation method was utilized in the development of PLS-DA calibration models [35].

2.2.3. Variable Selection. Five different methods of selecting variables were tested to see which of these methods may produce more accurate prediction results. These methods include backward variable elimination (BVE), genetic algorithm (GA), uninformative variable elimination (UVE), and subwindow permutation analysis (SwPA).

SwPA. The SwPA, when paired with the PLS-DA model, has the potential to make the model more effective and faster for analyzing large datasets. This is because the SwPA offers the influence of each variable individually, without taking into account the influence of the other factors. Additional information can be found in the reports that Mehmood and his coworkers [36] as well as Li and his coworkers [38] published.

IPW. The IPW variable selection was introduced by Forina and coworkers [39]. The method is predicated on the PLS model of each predictor's effect on the response, and it iteratively changes the original X-variables to eliminate the variables that are of the least importance. In the field of spectrometry, successful use of this method has been accomplished in the past [40].

BVE. Backward variable elimination was first ascribed by Frank for the elimination of noninformative variables [41]. Later, in an upgraded version, it was utilized for wavelength selection [42]. The method works by first sorting the variables using a filter measurement and then using a threshold to eliminate a subset of the least informative variables. This process is continued until there is no longer a need for any more elimination.

GA. The GA, which is derived from the concepts of genetics and natural selection, has developed into a tool for optimization that conducts a search that is both random and global inside a space that has a high dimension. By sampling a broad parameter space at each

stage of the optimization, GA might escape local optima and find global optima in a relatively short time. It has been extensively utilized for variable selection in multivariate spectroscopic calibration [43]. The steps of the genetic algorithm are explained in the study published by Mehmood and colleagues [36].

UVE. Before employing the PLS model, the UVE procedures that have been developed by Centner and coworkers with PLS models included the addition of artificial noise variables to the predictor set [44]. It does away with the habitat variables that are of lesser value compared to the artificial noise variables. This process is performed repeatedly until a satisfactory model is acquired.

3. Results and Discussion

3.1. NIRs Spectra. Figure 3 depicts the average of the NIRs absorbance spectra of raw Chinese quinces fruits grown in six different habitats. The raw fruit spectra show that all of the spectra have a relatively similar shape, and there is only a little amount of variation between the spectra of each habitat. However, after going through the first derivative preprocessing step, the raw fruit spectra showed that there were some major disparities across the different habitat groups. There were two strong bands of water absorbance at 1450 and 1950 nm that were connected to the overtone of -OH bands. The $-CH_3$ groups, such as methyl, methylene, and ethylene, were responsible for the peaks that appear at around 1250 nm, 1700 nm, 2000 nm, and 2150 nm, respectively [14, 45, 46]. In Figure 3(a), the observed spectra consisted of two distinct peaks and one broad peak, resulting in a total of three spectra. Conversely, Figure 3(b) exhibits a total of five spectra. Specifically, the absorption peak observed at a wavelength of 2,270 nm was attributed to the vibrational modes of CH-stretch and CH-deformation combination originating from the $-CH_3$ moiety of ethanol [47, 48]. Likewise, the absorption peak observed at approximately 2,300 nm is plausibly linked to the $-CH_2$ functional group present in ethanol [47–51]. The NIRs region ranging from 1,650 to 1,750 nm is associated with the first overtones of the CH-stretch in both $-CH_3$ and $-CH_2$ functional groups [47–51]. Additional research has demonstrated that methanol-based solutions containing phenolic compounds and tannins exhibit comparable absorption patterns within these specified regions, despite variations in concentration [52]. This is particularly relevant to the spectral regions centered at 1,650 and 1,850 nm, as well as the range between 2,100 and 2,300 nm. Within this range, a prominent absorption characteristic associated with tannins has been identified at approximately 2,140 nm [52]. Therefore, it is possible that the observed alterations in this region reflect differences in concentrations of sugar, ethanol, phenolics, and tannins.

The PLS-DA models' sensitivity in the classification of the six different habitats attained the best results from the first derivative spectra for both the calibration and validation sets. The correct classification specificity for the calibration set was 91%, while it was 95% for the validation set. For this reason, the optimal wavenumber selection was achieved by the application of the first derivative preprocessing approach.

3.2. Variable's Selection. PLS-DA was used in conjunction with the various variable selection methods to develop the final model. Table 2 illustrates the specificity of the PLS-DA models for both the calibration and validation sets for each variable selection method. The UVE variable selection approach achieved higher specificity for the calibration and validation sets, with scores of 0.93 and 0.98, respectively. This resulted in the best classification specificity that was achieved after employing this method. When compared to PLS-DA with no variable selection, which utilized 256 variables and 8 factors, the number of variables was decreased from 256 to 70 with the usage of UVE, and the number of PLS factors was lowered from 8 to 7. The specificity of BVE's classification was the least and came in at 0.89 for the calibration set and 0.93 for the validation set, respectively. Except for the GA method, which only eliminated 14 variables from the habitual spectrum, the other variable selection methods did not increase the classification model specificity, despite the fact that the number of variables was significantly decreased.

One notable advantage of the UVE-PLS method, in comparison to alternative variable selection methods, is its user independence, which eliminates any potential configuration issues [44]. In their study, Koshoubu et al. [53] presented an adapted iteration of UVE-PLS, wherein they incorporated the prediction error sum of squares. This modification was employed to exclude uninformative samples, considering both wavelength variables and concentration variables [54]. The UVE-PLS method is utilized to identify the wavelength variables that contain relevant information based on the regression coefficients obtained from PLS modeling. The coefficients of the PLS regression are acquired using the leave-one-out technique on the calibration samples. Nevertheless, the leave-one-out method presents a compelling issue. As highlighted by Martens and Dardenne [55], the leave-one-out technique employed in multivariate data analysis typically tends to overfit on average, resulting in an underestimation of the actual predictive error. Hence, the incorporation of the leave-one-out method in the UVE-PLS algorithm introduces the aforementioned drawbacks, potentially resulting in the overfitting of the prediction model.

Table 3 provides an overview of the correct classification percentages for both the calibration set and the validation set both before and after the application of UVE. Overall, PLS-DA-UVE produced optimal results when used for the classification of the different habitats of quince fruits. PLS-DA-UVE was superior to PLS-DA in terms of improving the specificity of classification for Anhui, Shandong, and Yunnan in the calibration set when compared to PLS-DA with no variable selection. The specificity of the Chongqing and Zhejiang habitats remained the same, whereas it decreased for the Hubei habitat. Using the UVE method in conjunction with PLS-DA resulted in a classification

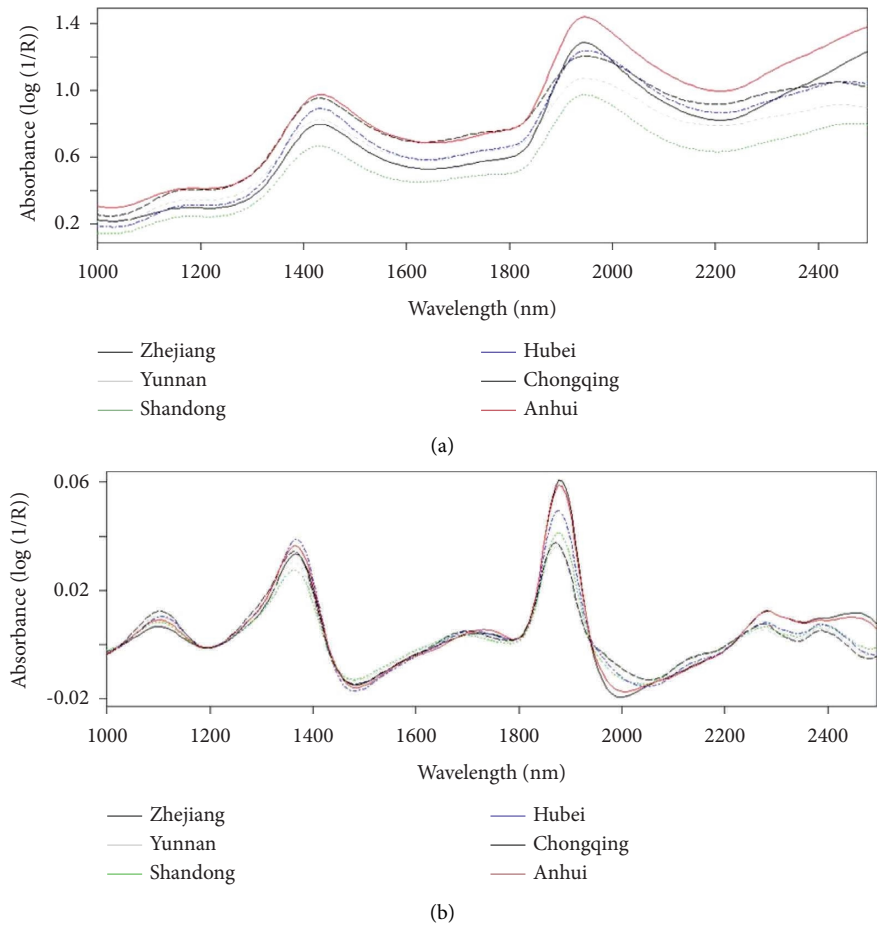


Figure 3: NIRs spectra of the Chinese quince fruits from six different habitats: (a) raw fruit and (b) first derivative preprocessing.

Table 2: Results of the classification of Chinese quince fruits from six different habitats using the PLS-DA full-wavelength and variable selection methods, respectively.

Model	Calibration		Validation		Variables	PLS factors
	Accuracy	Error	Accuracy	Error		
PLS-DA	0.91	0.09	0.95	0.05	256	8
PLS-DA-SwPA	0.91	0.09	0.93	0.07	210	8
PLS-DA-IPW	0.91	0.09	0.93	0.07	106	5
PLS-DA-BVE	0.89	0.11	0.93	0.07	101	10
PLS-DA-GA	0.91	0.09	0.95	0.05	242	8
PLS-DA-UVE	0.93	0.07	0.98	0.02	70	7

specificity of 100% achieved in the validation set for quinces belonging to the regions of Anhui, Chongqing, Hubei, Shandong, and Zhejiang. The specificity of the classification of quince fruit harvested in Yunnan habitats improved only marginally, ranging from 86% to 88%. PLS-DA-UVE succeeded in achieving the best overall performance, indicating the superiority of this method over others, it effectively classifies the habitat of Chinese quince fruits using NIRs spectral data.. It was found that using UVE in conjunction with PLS-DA methods might produce a result that was more reliable and specific [56]. A similar result was observed when combining UVE with PLS-DA to determine the linoleic acid concentration in eight different types of edible vegetable oils [57]. This indicates that the FT-IR transmission spectroscopy approach combined with the UVE method is promising for the quick detection of glycerol monolaurate [58].

Figure 4 presents the PLS-DA and PLS-DA-UVE score plots for factors 1 and 2, respectively. The PLS-DA score plot (Figure 4(a)) shows that the fruits from each of the six habitats may be distinguished from one another. This might be because Chinese quinces grow in a wide variety of habitats, each of which is unique in terms of the soil, climate, and growing conditions, even though there is some commonality. It is evident from observing Figure 4(b) that the six

Table 3: Percentage of correct classification of Chinese quince fruits from six different habitats using the PLS-DA full-wavelength and uninformative variable elimination (UVE) methods, respectively.

Habitats	PLS-DA		PLS-DA-UVE	
	Calibration	Validation	Calibration	Validation
Anhui	93	100	95	100
Chongqing	94	100	94	100
Hubei	93	100	90	100
Shandong	81	75	82	100
Yunnan	83	86	94	88
Zhejiang	100	100	100	100

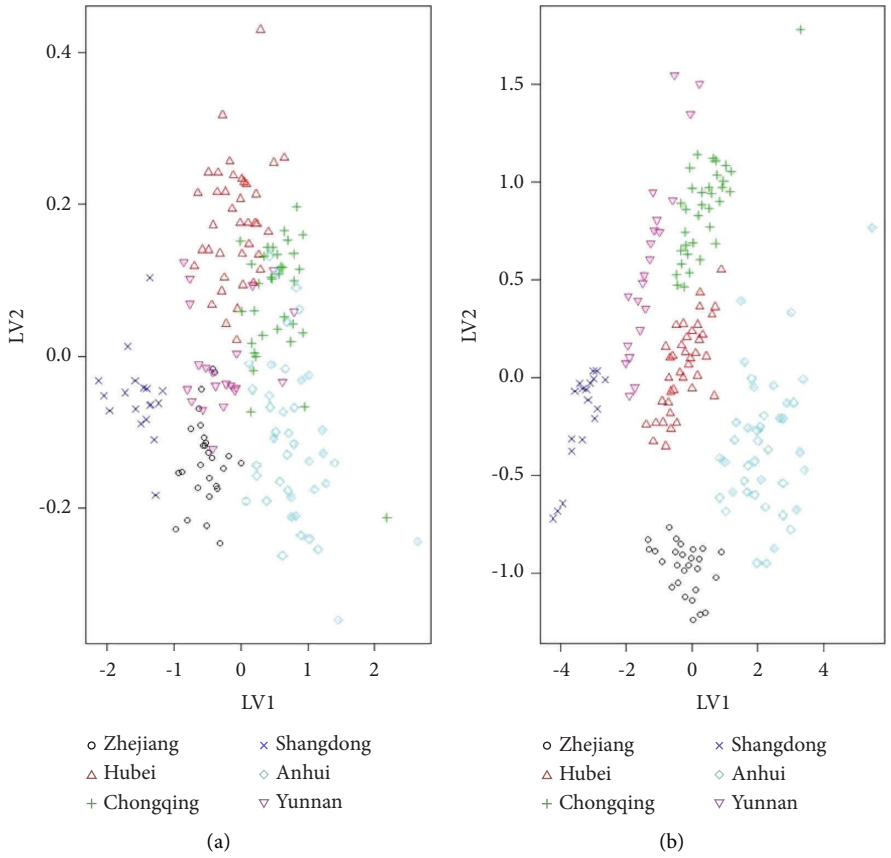


Figure 4: Score plots of Chinese quince fruits from six different habitats: (a) PLS-DA and (b) PLS-DA-UVE.

clusters have been successfully differentiated using UVE in conjunction with PLS-DA.

4. Conclusion

The NIRs technique was employed in this study to successfully classify samples of Chinese quince fruit, resulting in significant disparities observed among the habitat groups obtained from six different habitats. Raw fruit spectra in the range of 1000 to 2500 nm were found when PLS-DA models were combined with the first derivative preprocessing method. This has the potential to be employed as a fast and nondestructive method for differentiating the habitat of Chinese quinces. Following an examination of several other variable selection methods, the study found that the UVE variable selection method, when used in conjunction with the PLS-DA method, produces more accurate classifications for the six different habitats. In addition, the findings suggested that the discrimination against the habitat of Chinese quinces can be due to the difference in the chemical composition of Chinese quince fruits, which resulted from the different climatic and geographical conditions of the habitat in which Chinese quinces were grown. This difference in the chemical composition of Chinese quince fruits was caused by the fact that Chinese quinces were grown in a habitat in which they had to adapt to different conditions. In addition, the findings of the study suggest that PLS-DA can be used as an alternative method for classifying the habitats of Chinese quince fruits. This will help in identifying the primary factors that cause significant variation in the habitats, composition, and quality of Chinese quince fruits when combined with other methods like polynomial

multivariate and multiregression analysis. Furthermore, the focus of future work proposes combining near-infrared spectroscopy with other methods of stoichiometry in which the products and reactants are compared, and the Law of Conservation of Mass and Energy is applied to get quantitative information on the reaction, which can be utilized to further investigate the main factors impacting the variation of Chinese quince fruits in different habitats. Therefore, it can be asserted that the current investigation possesses notable strengths and limitations, along with implications for subsequent research endeavors and/or clinical applications. The current study's strengths can be inferred from the utilization of NIRs, a noninvasive and expeditious analytical technique that offers valuable insights into the composition and characteristics of the samples. The current investigation places its emphasis on the classification of Chinese quince fruits originating from various habitats. This classification process has the potential to contribute to the enhancement of quality control, grading, and sorting procedures for these fruits. In the current investigation, NIRs is employed to nondestructively analyze samples, rendering it an invaluable instrument for evaluating the quality of fruits while preserving their usability and market value. NIRs is recognized for its rapid analysis capabilities in the current investigation, providing a distinct advantage for time-sensitive applications such as quality control in fruit processing. Moreover, the current study exhibits certain limitations. The current investigation may exhibit a constrained sample size, potentially impacting the extent to which the findings can be extrapolated. Another limitation is the potential absence of external validation through the utilization of independent datasets or samples from diverse geographical locations, which could enhance the credibility of the classification models. The present study suggests that there are important implications for future research and clinical practice. Specifically, it is recommended that future research endeavors focus on validating the findings using larger and more diverse samples. This approach will help to improve the reliability and generalizability of the classification models. To ascertain pivotal spectral characteristics, forthcoming investigations should prioritize the identification of distinct spectral attributes linked to the categorization of Chinese quince fruits. This can facilitate comprehension of the inherent chemical composition and qualitative characteristics of these fruits. Potential applications in clinical practice encompass the utilization of NIRs for expedited categorization of fruits. This methodology can be further extrapolated to diverse domains, including the determination of the caliber and genuineness of medicinal plants as well as the evaluation of the nutritional constitution of food products. It may also have implications in clinical practice, such as the expedited identification of diseases or conditions through the analysis of spectral signatures in biological samples. The incorporation of additional analytical methodologies can result in a synergistic effect, whereby the combination of NIRs with other analytical techniques facilitates the acquisition of a more exhaustive and precise dataset. Future investigations may delve into the synergistic combination of NIRs with complementary techniques, such as chromatography or mass spectrometry, in order to augment the analytical capabilities pertaining to Chinese quince fruits or analogous specimens. In a nutshell, this study showcases the capacity of NIRs for categorizing and evaluating the quality of Chinese quince fruits. Subsequent investigations can expand upon these results to investigate wider applications and enhance the technique's efficiency.

Data Availability

Data for this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Science and Technology Basic Resources Survey Program of China (2019FY100803-02), the Central Finance Forest and Grass Science and Technology Demonstration Project (GTH[2024]2), and Fundamental Research Funds of CAF (CAFYBB2018SY016).

References

[1] M. M. Wang, H. B. Chen, J. H. Wang, and S. Li, "Genetic relationship of *Chaenomeles* cultivars revealed by SRAP analysis," *Scientia Agricultura Sinica*, vol. 43, no. 3, pp. 542–551, 2010.

[2] W. Shao, Y. Li, S. Diao, J. Jiang, and R. Dong, "Rapid classification of Chinese quince (*Chaenomeles speciosa* Nakai) fruit habitat by near-infrared spectroscopy and multivariate calibration," *Analytical and Bioanalytical Chemistry*, vol. 409, no. 1, pp. 115–120, 2017.

[3] K. Patel, V. Kumar, M. Rahman, A. Verma, and D. K. Patel, "New insights into the medicinal importance, physiological functions and bioanalytical aspects of an important bioactive compound of foods 'Hyperin': health benefits of the past, the present, the future," *Beni-Suef University Journal of Basic and Applied Sciences*, vol. 7, no. 1, pp. 31–42, 2018.

[4] S. Y. Zhang, L. Y. Han, H. Zhang, and H. L. Xin, "*Chaenomeles speciosa*: a review of chemistry and pharmacology," *Biomedical Reports*, vol. 2, no. 1, pp. 12–18, 2014.

[5] A. F. El Sheikha, I. Métayer, and D. Montet, "A biological bar code for determining the geographical origin of fruit by using 28S rDNA fingerprinting of fungal communities by PCR-DGGE: an application to physalis fruits from Egypt," *Food Biotechnology*, vol. 25, no. 2, pp. 115–129, 2011.

[6] F. Licciardello, G. Muratore, C. Avola, and E. Maccarone, "Geographical habitat assessment of orange juices by comparison of free aminoacids distribution," *International Symposium on Citrus Biotechnology*, vol. 892, pp. 389–394, 2009.

[7] S. Nojima, C. Linn Jr, and W. Roelofs, "Identification of host fruit volatiles from flowering dogwood (*Cornus florida*) attractive to dogwood-habitat *Rhagoletis pomonella* flies," *Journal of Chemical Ecology*, vol. 29, no. 10, pp. 2347–2357, 2003.

[8] Y. Liu, X. Chen, and A. Ouyang, "Nondestructive determination of pear internal quality indices by visible and

near-infrared spectrometry," *LWT--Food Science and Technology*, vol. 41, no. 9, pp. 1720–1725, 2008.

[9] B. Hasanzadeh, Y. Abbaspour-Gilandeh, A. Soltani-Nazarloo, E. D. L. Cruz-Gámez, J. L. Hernández-Hernández, and M. Martínez-Arroyo, "Non-destructive measurement of quality parameters of apple fruit by using visible/near-infrared spectroscopy and multivariate regression analysis," *Sustainability*, vol. 14, no. 22, Article ID 14918, 2022.

[10] J. Ugarte Fajardo, O. Bayona Andrade, R. Criollo Bonilla et al., "Early detection of black Sigatoka in banana leaves using hyperspectral images," *Applications in Plant Sciences*, vol. 8, no. 8, Article ID e11383, 2020.

[11] X. Fu, Y. Ying, Y. Zhou, L. J. Xie, and H. R. Xu, "Application of NIR spectroscopy for firmness evaluation of peaches," *Journal of Zhejiang University Science B*, vol. 9, no. 7, pp. 552–557, 2008.

[12] A. Moghimi, M. H. Aghkhani, A. Sazgarnia, and M. Sarmad, "Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit," *Biosystems Engineering*, vol. 106, no. 3, pp. 295–302, 2010.

[13] A. M. Cavaco, P. Pinto, M. D. Antunes, J. M. D. Silva, and R. Guerra, "Rocha'pear firmness predicted by a Vis/NIR segmented model," *Postharvest Biology and Technology*, vol. 51, no. 3, pp. 311–319, 2009.

[14] K. B. D. S. Lobato, P. D. Alamar, E. T. D. S. Caramês, and J. A. L. Pallone, "Authenticity of freeze-dried açai pulp by near-infrared spectroscopy," *Journal of Food Engineering*, vol. 224, pp. 105–111, 2018.

[15] D. Eisenstecken, B. Stürz, P. Robatscher, L. Lozano, A. Zanella, and M. Oberhuber, "The potential of near infrared spectroscopy (NIRS) to trace apple origin: study on different cultivars and orchard elevations," *Postharvest Biology and Technology*, vol. 147, pp. 123–131, 2019.

[16] X. Li, J. Huang, Y. Xiong, J. Zhou, X. Tan, and B. Zhang, "Determination of soluble solid content in multi-origin 'Fuji' apples by using FT-NIR spectroscopy and an origin discriminant strategy," *Computers and Electronics in Agriculture*, vol. 155, pp. 23–31, 2018.

[17] J. Vincent, H. Wang, O. Nibouche, and P. Maguire, "Differentiation of apple varieties and investigation of organic status using portable visible range reflectance spectroscopy," *Sensors*, vol. 18, no. 6, p. 1708, 2018.

[18] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.

[19] R. G. Brereton and G. R. Lloyd, "Partial least squares discriminant analysis: taking the magic away," *Journal of Chemometrics*, vol. 28, no. 4, pp. 213–225, 2014.

[20] K. M. Oksman-Caldentey and D. Inzé, "Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites," *Trends in Plant Science*, vol. 9, no. 9, pp. 433–440, 2004.

[21] A. L. Boulesteix and K. Strimmer, "Partial least squares: a versatile tool for the analysis of high-dimensional genomic data," *Briefings in Bioinformatics*, vol. 8, no. 1, pp. 32–44, 2006.

[22] G. Blekherman, R. Laubenbacher, D. F. Cortes et al., "Bioinformatics tools for cancer metabolomics," *Metabolomics*, vol. 7, no. 3, pp. 329–343, 2011.

[23] E. Szymanska, E. Saccenti, A. K. Smilde, and J. A. Westerhuis, "Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies," *Metabolomics*, vol. 8, no. 1, pp. S3–S16, 2012.

[24] C. Christin, H. C. Hoefsloot, A. K. Smilde et al., "A critical assessment of feature selection methods for biomarker discovery in clinical proteomics," *Molecular & Cellular Proteomics*, vol. 12, no. 1, pp. 263–276, 2013.

[25] J. L. Izquierdo-García, I. Rodríguez, A. Kyriazis et al., "A novel R-package graphic user interface for the analysis of metabonomic profiles," *BMC Bioinformatics*, vol. 10, no. 1, pp. 363–410, 2009.

[26] K. A. Le Cao, I. Gonzalez, and S. Dejean, "integrOmics: an R package to unravel relationships between two omics datasets," *Bioinformatics*, vol. 25, no. 21, pp. 2855-2856, 2009.

[27] T. Wang, K. Shao, Q. Chu et al., "Automics: an integrated platform for NMR-based metabonomics spectral processing and data analysis," *BMC Bioinformatics*, vol. 10, no. 1, p. 83, 2009.

[28] E. Want and P. Masson, "Processing and analysis of GC/LC-MS-based metabolomics data," *Methods in Molecular Biology*, vol. 708, pp. 277–298, 2011.

[29] J. Xia, N. Psychogios, N. Young, and D. S. Wishart, "MetaboAnalyst: a web server for metabolomic data analysis and interpretation," *Nucleic Acids Research*, vol. 37, pp. W652–W660, 2009.

[30] G. Quintás, N. Portillo, J. C. García-Cañaveras, J. V. Castell, A. Ferrer, and A. Lahoz, "Chemometric approaches to improve PLSDA model outcome for predicting human non-alcoholic fatty liver disease using UPLC-MS as a metabolic profiling tool," *Metabolomics*, vol. 8, no. 1, pp. 86–98, 2012.

[31] M. Bassbasi, M. De Luca, G. Ioele, A. Oussama, and G. Ragno, "Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data," *Journal of Food Composition and Analysis*, vol. 33, no. 2, pp. 210–215, 2014.

[32] T. Pholpho, S. Pathaveerat, and P. Sirisomboon, "Classification of longan fruit bruising using visible spectroscopy," *Journal of Food Engineering*, vol. 104, no. 1, pp. 169–172, 2011.

[33] A. A. F. Zielinski, C. W. I. Haminiuk, C. A. Nunes, E. Schnitzler, S. M. van Ruth, and D. Granato, "Chemical composition, sensory properties, provenance, and bioactivity of fruit juices as assessed by chemometrics: a critical review and guideline," *Comprehensive Reviews in Food Science and Food Safety*, vol. 13, no. 3, pp. 300–316, 2014.

[34] Y. Zhao, Z. Zhao, P. Shan, S. Peng, J. Yu, and S. Gao, "Calibration transfer based on affine invariance for NIR without transfer standards," *Molecules*, vol. 24, no. 9, p. 1802, 2019.

[35] L. Xu, W. Sun, C. Wu, Y. Ma, and Z. Chao, "Discrimination of *Trichosanthis* fructus from different geographical habitats using near infrared spectroscopy coupled with chemometric techniques," *Molecules*, vol. 24, no. 8, p. 1550, 2019.

[36] T. Mehmood, K. H. Liland, L. Snipen, and S. Sæbø, "A review of variable selection methods in partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 118, pp. 62–69, 2012.

[37] R. C. Team, "R: a language and environment for statistical computing," 2013, https://www.r-project.org/.

[38] H. D. Li, Y. Z. Liang, D. S. Cao, and Q. S. Xu, "Model-population analysis and its applications in chemical and biological modeling," *Trends in Analytical Chemistry*, vol. 38, pp. 154–162, 2012.

[39] M. Forina, C. Casolino, and C. Pizarro Millan, "Iterative predictor weighting (IPW) PLS: a technique for the elimination of useless predictors in regression problems," *Journal of Chemometrics*, vol. 13, no. 2, pp. 165–184, 1999.

[40] D. Chen, B. Hu, X. Shao, and Q. Su, "Variable selection by modified IPW (iterative predictor weighting)-PLS (partial

least squares) in continuous wavelet regression models," *Analyst*, vol. 129, no. 7, pp. 664–669, 2004.

[41] I. E. Frank, "Intermediate least squares regression method," *Chemometrics and Intelligent Laboratory Systems*, vol. 1, no. 3, pp. 233–242, 1987.

[42] J. A. Fernández Pierna, O. Abbas, V. Baeten, and P. Dardenne, "A backward variable selection method for PLS regression (BVSPLS)," *Analytica Chimica Acta*, vol. 642, no. 1-2, pp. 89–93, 2009.

[43] B. M. Smith and P. J. Gemperline, "Wavelength selection and optimization of pattern recognition methods using the genetic algorithm," *Analytica Chimica Acta*, vol. 423, no. 2, pp. 167–177, 2000.

[44] V. Centner, D. L. Massart, O. E. de Noord, S. de Jong, B. M. Vandeginste, and C. Sterna, "Elimination of uninformative variables for multivariate calibration," *Analytical Chemistry*, vol. 68, no. 21, pp. 3851–3858, 1996.

[45] S. Diao, W. Shao, Q. Luan, R. Dong, and J. Jiang, "Estimation of pericarp saponin concent in *Sapindus mukorossi* by using near infrared reflectance spectroscopy," *Chemistry & Industry of Forest Products*, vol. 34, no. 5, pp. 91–96, 2014.

[46] N. Sinelli, L. Cerretani, V. D. Egidio, A. Bendini, and E. Casiraghi, "Application of near (NIR) infrared and mid (MIR) infrared spectroscopy as a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity," *Food Research International*, vol. 43, no. 1, pp. 369–375, 2010.

[47] R. G. Dambergs, A. Kambouris, I. L. Francis, and M. Gishen, "Rapid analysis of methanol in grape-derived distillation products using near-infrared transmission spectroscopy," *Journal of Agricultural and Food Chemistry*, vol. 50, no. 11, pp. 3079–3084, 2002.

[48] D. Cozzolino, M. J. Kwiatkowski, M. Parker et al., "Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy," *Analytica Chimica Acta*, vol. 513, no. 1, pp. 73–80, 2004.

[49] I. Murray, "NIR/NIT conference," in *Proceedings of the International NIR/NIT Conference*, J. Hollo, K. J. Kaffka, and J. L. Gonczy, Eds., Akademiai Kiado, Budapest, Hungary, 1986.

[50] B. G. Osborne, T. Fearn, and P. H. Hindle, "Practical NIR spectroscopy with applications in food and beverage analysis," *Longman scientific and technical*, vol. 4, p. 227, 1993.

[51] D. Cozzolino, H. E. Smyth, and M. Gishen, "Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins," *Journal of Agricultural and Food Chemistry*, vol. 51, no. 26, pp. 7703–7708, 2003.

[52] J. Soukupova, B. N. Rock, and J. Albrechtova, "Spectral characteristics of lignin and soluble phenolics in the near infrared-a comparative study," *International Journal of Remote Sensing*, vol. 23, no. 15, pp. 3039–3055, 2002.

[53] J. Koshoubu, T. Iwata, and S. Minami, "Application of the modified UVE-PLS method for a mid-infrared absorption spectral data set of water—ethanol mixtures," *Applied Spectroscopy*, vol. 54, no. 1, pp. 148–152, 2000.

[54] J. Koshoubu, T. Iwata, and S. Minami, "Elimination of the uninformative calibration sample subset in the modified UVE (uninformative variable elimination)–PLS (partial least squares) method," *Analytical Sciences*, vol. 17, no. 2, pp. 319–322, 2001.

[55] H. A. Martens and P. Dardenne, "Validation and verification of regression in small data sets," *Chemometrics and Intelligent Laboratory Systems*, vol. 44, no. 1-2, pp. 99–121, 1998.

[56] W. Cai, Y. Li, and X. Shao, "A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra," *Chemometrics And Intelligent Laboratory Systems*, vol. 90, no. 2, pp. 188–194, 2008.

[57] D. Wu, X. Chen, P. Shi, S. Wang, F. Feng, and Y. He, "Determination of α-linolenic acid and linoleic acid in edible oils using near-infrared spectroscopy improved by wavelet transform and uninformative variable elimination," *Analytica Chimica Acta*, vol. 634, no. 2, pp. 166–171, 2009.

[58] X. Chen, D. Wu, Y. He, and S. Liu, "Detecting the quality of glycerol monolaurate: a method for using Fourier transform infrared spectroscopy with wavelet transform and modified uninformative variable elimination," *Analytica Chimica Acta*, vol. 638, no. 1, pp. 16–22, 2009.