View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Crossref

Hindawi Publishing Corporation
Journal of Spectroscopy
Volume 2016, Article ID 7927286, 7 pages
http://dx.doi.org/10.1155/2016/7927286

Research Article

Mid-Infrared Spectroscopy for Coffee Variety Identification: Comparison of Pattern Recognition Methods

Chu Zhang, Chang Wang, Fei Liu, and Yong He

College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310038, China

Correspondence should be addressed to Fei Liu; fliu@zju.edu.cn and Yong He; yhe@zju.edu.cn

Received 10 October 2015; Revised 29 December 2015; Accepted 30 December 2015

Academic Editor: Eugen Culea

Copyright © 2016 Chu Zhang 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 potential of using mid-infrared transmittance spectroscopy combined with pattern recognition algorithm to identify coffee variety was investigated. Four coffee varieties in China were studied, including Typica Arabica coffee from Yunnan Province, Catimor Arabica coffee from Yunnan Province, Fushan Robusta coffee from Hainan Province, and Xinglong Robusta coffee from Hainan Province. Ten different pattern recognition methods were applied on the optimal wavenumbers selected by principal component analysis loadings. These methods were classified as highly effective methods (soft independent modelling of class analogy, support vector machine, back propagation neural network, radial basis function neural network, extreme learning machine, and relevance vector machine), methods of medium effectiveness (partial least squares-discrimination analysis, K nearest neighbors, and random forest), and methods of low effectiveness (Naive Bayes classifier) according to the classification accuracy for coffee variety identification.

1. Introduction

Coffee is one of the most important and popular beverages all over the world. Coffee plants are cultivated in over 70 countries. Coffee trade and consumption are the important income source for many people and many countries [1]. Because of the vast plant territory and varieties of coffee plants, the quality of coffee beans varies and is significantly related to the growth conditions and the varieties. Identification of coffee bean varieties is crucial for coffee trade and consumption.

Traditional analytical methods for identifying coffee bean are laboratory-based, costly, time-consuming, and requiring technical skills [2–6]. Therefore, simpler, faster, and cheaper methods for coffee bean varieties determination are required. In recent years, spectroscopy techniques have been reported as simple, fast, and reliable methods in different fields. Studies have been reported for identification of coffee varieties, coffee origin, and coffee blends using near-infrared spectroscopy [7–9], mid-infrared spectroscopy [10–12], nuclear magnetic resonance spectroscopy [13–15], and Raman spectroscopy [16–18]. These studies obtained good performance, indicating the feasibility of using spectroscopy techniques in coffee industries.

Among these methods, mid-infrared spectroscopy ($400\,cm^{-1}$–$4000\,cm^{-1}$) is used to study the fundamental vibrations and associated rotational-vibrational structure of chemical bonds [19]. Mid-infrared spectroscopy has been used as an efficient analytical tool in various fields, such as agriculture [20], food [21], oil [22], medical [23], textile [24], and pharmaceutical [19, 25]. Many papers have been published for identification or classification of sample varieties and cultivars [26–28] as well as coffee varieties and coffee blends [10–12] with the aid of multivariate analysis methods.

Since identification of coffee variety by mid-infrared spectroscopy has been proved to be feasible, the methods to build classification models for better and robust classification results should be further studied. Many pattern recognition methods have been applied for spectral data analysis of classification issues, especially the supervised methods for constructing the classification models. Different pattern recognition methods showed different results due to different algorithm principles. In many studies, at least 3 methods were

used and compared for analysis. In this study, we applied 10 pattern recognition methods for coffee variety identification to select optimal recognition methods for practical application, including partial least squares-discrimination analysis (PLS-DA) [29], *K* nearest neighbors (*K*NN) [30], SIMCA [31], support vector machine (SVM) [32], back propagation neural network (BPNN) [33], radial basis function neural network (RBFNN) [34], extreme learning machine (ELM) [35], random forest (RF) [36], Naive Bayes classifier [37], and relevance vector machine (RVM) [38]. Among these methods, PLS-DA, *K*NN, SIMCA, SVM, and BPNN are the most used methods in spectral data analysis.

The main objective of this study was to use mid-infrared spectroscopy for coffee bean variety identification with different pattern recognition methods. The specific objectives were to (1) select the optimal wavenumbers which contributed most to the identification to coffee bean varieties; (2) build classification models by using 10 different pattern recognition methods; (3) compare and select the most effective pattern recognition methods for coffee bean variety identification.

2. Material and Methods

2.1. Sample Preparation. Four varieties of coffee beans in China (Typica Arabica coffee from Yunnan Province, Catimor Arabica coffee from Yunnan Province, Fushan Robusta coffee from Hainan Province, and Xinglong Robusta coffee from Hainan Province) were collected. All coffee beans were collected at the same year and medium toasted. Six hundred coffee beans of each variety were collected and stored in a vacuum glass box. Twenty coffee beans of each variety were grounded, screened by 80 mesh sieve, and dried as one sample, and 30 samples of each variety were prepared. The samples of each variety were randomly divided into the training set and the test set with the ratio of 2 : 1 (20 samples of each variety for training and 10 samples of each variety for test).

2.2. Mid-Infrared Spectra Collection. The mid-infrared spectra of sample were acquired by a Jasco FT/IR-4100 spectrometer (Japan) in the spectral range of 400 cm^{-1}–4000 cm^{-1}. 20 mg of each sample was mixed with 980 mg KBr powders, the mixture was grounded and mixed thoroughly. The mixture was put into the tablet machine for tableting, and the sample tablet was used for transmittance MIR spectral data collection. For each sample, 32 times of scans were applied with the resolution of 4 cm^{-1} and the average spectrum of the 32 spectra was used as the transmittance spectrum of the sample.

2.3. Multivariate Analysis Methods. Principal component analysis (PCA) is the most widely used method for qualitative analysis of spectral data. PCA linearly transforms the original data into new variables (principal components (PCs)) which are the linear combination of the original data. The first PC (PC1) has the direction of maximum variance, and the second PC (PC2) has the direction of second largest variance, and so do the rest PCs. Generally, the first few PCs could explain the most variance of the original data. In many cases, the 2D scores scatter plot by PC1 and PC2 was used to present the sample distributions, especially for the classification issues [39, 40].

PLS-DA is a supervised pattern recognition method based on PLSR. PLS-DA conducts PLSR with integers representing the categories as Y. The outputs of PLS-DA are real numbers with decimals. To determine which category the sample belongs to, a threshold value should be set. In this study, the threshold value was set as 0.5, indicating that the sample belongs to the category which was the nearest integer of the test value [29]. There are two approaches to be used in PLSR, PLS1 and PLS2. When the Y response consists only of 1 variable, PLS1 is applied. When there are more than one variable in Y response, PLS2 is used. In this study, only one variable was in Y response and PLS1-DA was used [41].

*K*NN is a supervised pattern recognition method. *K*NN calculates the distance between samples, and the category of the sample is determined by its k nearest neighbors (k samples with the smallest distance from the sample). The sample is classified by a majority vote of its k nearest neighbors [30].

SIMCA is a supervised pattern recognition method based on PCA scores. SIMCA firstly conducts PCA of each category and determines the optimal number of PCs for classification. The classification procedure is then implemented based on the PCA residuals [31].

SVM is a widely used machine learning method for regression and classification. SVM maps the original data into a high dimensional space and finds a hyperplane which has the largest distance to the nearest data point of any category. Then, the samples were classified. To conduct SVM, the kernel functions to be used are essential and significantly important [32].

BPNN is a widely used artificial neural network (ANN). BPNN uses error back propagation to modify the internal network weights after each training epoch until the goal of the training error or the training epochs of the network is achieved [33].

RBFNN is another widely used artificial neural network. RBFNN uses RBF as the activation function. RBFNN typically has three layers: an input layer, a hidden layer with a nonlinear RBF activation function, and a linear output layer. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters [34].

ELM is a single-hidden layer feedforward network (SLFN). ELM is a fast, simple method for regression and classification. In ELM, only the number of neurons needs to be set to obtain unique optimal solution [35].

RF is an ensemble method that uses a multitude of decision trees. RF constructs different decision trees, and the decision trees are independent of each other. To construct a random forest, the samples to train each decision are randomly selected from the training sample set by recovery sampling. The features to be used for the node of the decision tree are also randomly selected from the features from the training sample set. The output classification results are based on the output of each decision tree [36].

Naive Bayes classifier is a supervised pattern recognition method based on Bayes' theorem. It assumes that the features

are with strong independence. Naive Bayes classifier is a probabilistic classifier, and the features contribute independently to the probability of the sample category [37].

RVM is a machine learning method based on Bayesian inference. RVM is a special case of sparse Bayesian modelling. RVM has the same function form as SVM, and it has many common features as SVM. Unlike SVM, RVM provides probabilistic classification [38].

2.4. Optimal Wavenumber Selection. The number of the spectral data points of the acquired transmittance spectra of 400–4000 cm^{-1} was 3734 with the spectral resolution of 4 cm^{-1}. The large amount of data increase the computation time and the hardware level. Besides, the redundant information would result in complex, unstable, and inaccurate models. Optimal wavenumber (wavelength) selection is generally used to solve these problems. Optimal wavenumber (wavelength) selection methods work efficiently by selecting several wavenumbers (wavelengths) carrying most information from the original spectral data. In this study, PCA loadings were used to select the optimal wavenumbers. The peaks and valleys of loading plots of the first few PCs were selected as the optimal wavenumbers, and the selected wavenumbers contribute most to the loadings of each PC [42].

2.5. Model Evaluation and Software. The classification accuracy of the classification models was evaluated by the ratio of the number of correctly classified samples and the number of total samples (corrected classified rate) of the training set and the test set. The higher the classification accuracy was, the better performance the model obtained. *K*NN, SIMCA, SVM, BPNN, RBFNN, ELM, RF, Naive Bayes classifier, and RVM models were built on Matlab R 2010b (The Math Works, Natick, USA), and PLS-DA and PCA were conducted on Unscrambler® 10.1 (CAMO AS, Oslo, Norway).

3. Results and Discussion

3.1. Spectral Profiles. Considering the noises of the head and the end of the collected transmittance spectra in the range of 400–4000 cm^{-1}, only the spectra of 700–3600 cm^{-1} were used for analysis. The raw transmittance spectra are shown in Figure 1(a). It could be noticed that there were random noises of the spectral data. The preprocessing of the spectral data is necessary to reduce the noises. Wavelet transform (WT) is an efficient tool to remove the noises from the signals by a wavelet series with different spatial and frequency properties, and it has been used to remove the noises from the spectral data [43]. In this study, the wavelet function Daubechies 4 (db4) with the decomposition level 4 was applied after the trials of different wavelet function with different decomposition level.

The raw spectrum and the spectrum preprocessed by WT of a randomly selected sample are shown in Figure 1(a). It could be observed that the preprocessed spectra were much smoother than the unpreprocessed spectra without eliminating the critical transmittance peaks and valleys. It could be observed from the average spectra (Figure 1(b)) that the trend of the transmittance spectra of the 4 varieties was the same, while the transmittance values were quite different, showing obvious differences.

Table 1: The selected wavenumbers by principal component analysis loadings.

Method	Wavenumbers (cm^{-1})
PCA loading	835.0262, 911.2006, 1061.6211, 1268.9313, 1291.1086, 1322.9283, 1337.3918, 1374.9969, 1396.2101, 1398.1385, 1406.8167, 1418.3875, 1499.3831, 1511.9181, 1620.8765, 1633.4115, 1662.3385, 1741.4056, 1881.2195, 2315.1245, 2358.5151, 2359.4792, 2362.3721, 2927.4128, 3129.9019, 3148.2222, 3287.0718, 3479.9185, and 3546.4507

3.2. Principal Component Analysis. PCA was conducted on the preprocessed spectral data. The scores scatter plot of PC1 and PC2 is shown in Figure 2. PC1 and PC2 explained 91.275% and 4.159% of the total variance, respectively. It could be observed in Figure 2 that most of the samples could be distinguished from the samples of the other varieties, indicating the feasibility of coffee variety identification. Some observed overlaps indicated that further analysis for coffee varieties identification is needed.

3.3. Optimal Wavenumbers Selection. The first 4 PCs explained over 99.310% of the total variance. The loadings of the first 4 PCs were used to select the optimal wavenumbers. The peaks and valleys of the loading plot were selected (shown in Figure 3). In all, 29 optimal wavenumbers were selected, and the selected optimal wavenumbers are shown in Table 1.

The wavenumber near 1745 cm^{-1} (1741.4056 cm^{-1}) was characteristic of the C=O ester double bond of the triglyceride [44]. The selected wavenumbers (835.0262, 911.2006, and 1061.6211 cm^{-1}) were related to polysaccharides [45]. The selected wavenumbers (1268.9313, 1291.1086, 1322.9283, 1337.3918, 1374.9969, 1396.2101, 1398.1385, 1406.8167, 1418.3875, and 1499.3831 cm^{-1}) were related to organic acids and proteins [46]. The band around 2923 cm^{-1} (2927.4128 cm^{-1}) was characteristic of the length of the fatty acid chain [44].

3.4. Classification Models on Optimal Wavenumbers. Compared with the original data, the selection of optimal wavenumbers significantly reduced the number of input variables by 99.04%. The classification models were built on the optimal wavenumbers, and the results are shown in Table 2. To build classification models, the four varieties of coffee were assigned the category values of 1, 2, 3, and 4.

PLS-DA models were built with the spectral data as X and the category values as Y with leave-one-out cross validation. The threshold value of the PLS-DA model was set as 0.5. The optimal number of latent variables (LVs) was determined by the minimum Y residual variance. The classification accuracies of the training set and the test set were 86.25% and 80.00% with 7 LVs, indicating good classification results.

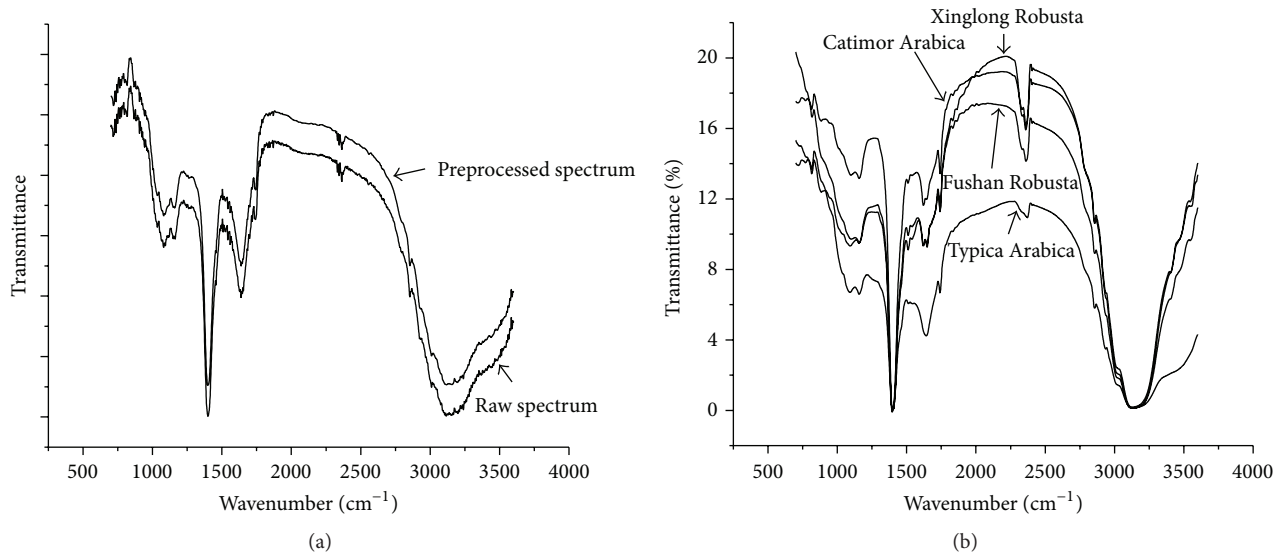


Figure 1: The raw mid-infrared spectrum and the mid-infrared spectrum preprocessed by wavelet transform of a randomly selected sample (a) and the average spectra of 4 varieties after preprocessing (b).

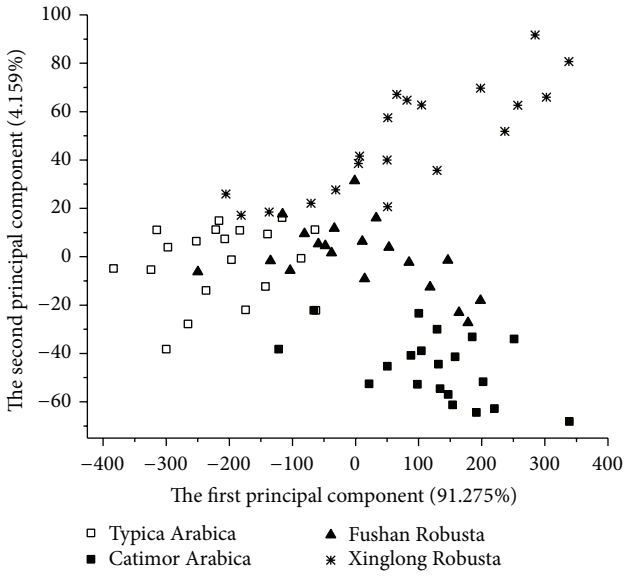


Figure 2: Scores scatter plot of the first principal component and the second principal component.

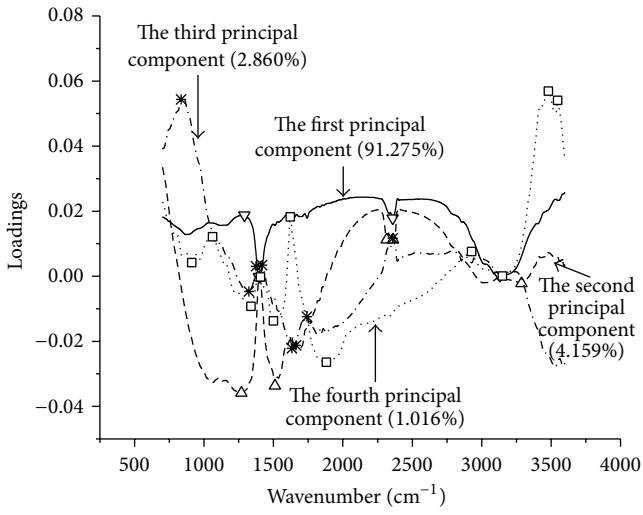


Figure 3: Wavenumbers selected by principal component analysis loadings (the marks (△, □, ∗, and ∇) referred to the selected wavenumbers of each loading).

The number of nearest neighbors was important for *K*NN model. To obtain optimal results, the number of nearest neighbors was set from 3 to 10. Euclidean distance was calculated as sample distances. The highest classification accuracy was achieved with 3 nearest neighbors. The classification accuracies of the training set and the test set were 95.00% and 90.00%, respectively.

For SIMCA model, different numbers of PCs were used and compared for each variety, and the best results were achieved on 10 PCs for each variety. The classification accuracies of the training set and the test set were 96.25% and 95.00%, respectively.

For BPNN model, the learning rate was 0.1 and the iteration epochs were 1000. The number of neurons of the hidden layer is determined by the following:

$$b = \sqrt{m+n} + a, \tag{1}$$

where m is the number of neurons in the input layer, n is the number of neurons in the output layer, and a is a constant between 1 and 10 [43]. After comparing the performances of BPNN models with different number of neurons in the hidden layer, the optimal number was determined as 7 with the classification accuracies of the training set and the test set of 100%.

Table 2: Classification results of the models on optimal wavenumbers.

	Training						Test					
	1	2	3	4		Total	1	2	3	4		Total
	Nr[a]/Nt[b]	Nr/Nt	Nr/Nt	Nr/Nt	Nr/Nt	Accuracy (%)	Nr/Nt	Nr/Nt	Nr/Nt	Nr/Nt	Nr/Nt	Accuracy (%)
SIMCA	20/20	20/20	17/20	20/20	77/80	96.25	10/10	10/10	8/10	10/10	38/40	95.00
SVM	20/20	20/20	20/20	20/20	80/80	100.00	10/10	10/10	9/10	9/10	38/40	95.00
ELM	20/20	20/20	20/20	20/20	80/80	100.00	9/10	10/10	10/10	10/10	39/40	97.50
BP	20/20	20/20	20/20	20/20	80/80	100.00	10/10	10/10	10/10	10/10	40/40	100.00
RBF	20/20	20/20	20/20	20/20	80/80	100.00	10/10	10/10	10/10	10/10	40/40	100.00
*K*NN	20/20	20/20	17/20	19/20	76/80	95.00	10/10	10/10	8/10	8/10	36/40	90.00
RF	20/20	20/20	20/20	20/20	80/80	100.00	10/10	10/10	8/10	9/10	37/40	92.50
PLS-DA	19/20	18/20	18/20	16/20	71/80	86.25	7/10	8/10	7/10	10/10	32/40	80.00
Naive Bayes classifier	19/20	17/20	13/20	9/20	58/80	72.50	10/10	10/10	4/10	5/10	29/40	72.50
RVM	20/20	20/20	20/20	20/20	80/80	100.00	10/10	10/10	9/10	9/10	38/40	95.00

a: Nr was the number of correctly classified samples; b: Nt was the total number of the samples.

The number of nodes in the hidden layer was important in ELM models. For ELM model, the optimal number of nodes in the hidden layer was based on a stepwise search. The number of nodes was selected from 1 to 80 with step of 1. The optimal classification accuracy was obtained by 52 nodes in the hidden layer. The classification accuracies of the training set and the test set were 100.00% and 97.50%, respectively.

For Naive Bayes classifier, the empirical prior probabilities for the classes were used. The classification accuracies of the training set and the test set were 72.50% and 72.50%, respectively.

For RBFNN model, the determination of spread value was important. In this study, the spread value was explored from 1 to 20, and the corresponding RBFNN model was built. The classification accuracies of the training set and the test set were both 100% with the spread value of 5.

For SVM, RBF was used as the kernel function. A grid search procedure was applied to search for the optimal penalty coefficient (C) and the kernel parameter (γ) of RBF. The classification accuracies of the training set and the test set were 100.00% and 95.00% with the optimal (C, γ) of (27.8576, 0.0068).

For RF model, the number of trees in the forest was set from 50 to 500, and the number of features to be used for each node was 5. The optimal number of trees was determined by the performances of RF models. The optimal classification accuracies of the training set and the test set were 100.00% and 92.50% with 50 trees.

For RVM model, the kernel function was selected as RBF, and the optimal kernel parameter was searched from 0.1 to 1. The RVM model obtained the classification accuracies of the training set and the test set of 100.00% and 95.00% with the optimal kernel parameter of 0.6.

It could be noted that the classification results of different models were different. The results of BPNN and RBFNN were excellent with classification accuracies of 100% in both the training set and the test set, while the results of Naive Bayes classifier were poor with classification accuracies lower than 80%. In general, the nonlinear classification models (SVM, BPNN, RBFNN, ELM, RF, and RVM) showed better results than the linear classification models (SIMCA, PLS-DA, *K*NN, and Naive Bayes classifier) in this study. The reason might be that the selected optimal wavenumbers contained more nonlinear features. According to the study of Balabin et al. [47], the classification models were divided into three categories by their classification accuracy: highly effective methods, methods of medium effectiveness, and methods of low effectiveness. In this study, the ten methods for coffee variety classification were divided into the above three categories by the classification accuracy. The methods with the classification accuracy over 95% in the training set and the test set were classified as highly effective methods, including SIMCA, SVM, ELM, BPNN, RBFNN, and RVM. The methods with the classification accuracy over 80% in the training set and the test set were classified as methods of medium effectiveness, including PLS-DA, *K*NN, and RF. The methods with the classification accuracy under 80% in the training set and the test set were classified as methods of low effectiveness, including Naive Bayes classifier. Moreover, all models were built on a computer with Intel Core i7 Processor and 16 GB memory, the computation time was less than 5 seconds, and the differences of computation time of all models were quite small.

As for the 4 coffee varieties, variety 3 (Fushan Robusta coffee from Hainan Province) and variety 4 (Xinglong Robusta coffee from Hainan Province) were more likely to be misclassified in all classification models, indicating the smaller differences between these two varieties.

The overall results indicated that the MIR spectroscopy with pattern recognition methods could efficiently identify the coffee varieties. The inputs of all models were significantly reduced from the original data, and the computation time of all the models showed no significant difference. The results showed that although ELM, RBFNN, and RVM were not frequently used in spectral data analysis, these methods could also be quite effective and promising for spectral data analysis and online application.

4. Conclusion

Mid-infrared spectroscopy combined with 9 different pattern recognition methods was successfully used to identify coffee varieties. The collected transmittance spectra were preprocessed by wavelet transform with db4 wavelet function and decomposition level of 4. The scores scatter plot of PCA showed the feasibility of identifying coffee varieties, and 29 optimal wavenumbers were selected by the loadings of the first 4 PCs. Ten classification models were built on the optimal wavenumbers. SIMCA, SVM, ELM, BPNN, RBFNN, and RVM models were classified as highly effective methods with classification accuracies over 95% in the training set and the test set; PLS-DA, *K*NN, and RF were classified as methods of medium effectiveness with the classification accuracy over 80% in the training set and the test set; Naive Bayes classifier was classified as methods of low effectiveness with classification accuracy lower than 80%. There was no significant difference of the computation time of different methods due to the optimal wavenumber selection. The highly effective methods were recommended for practical application. SVM, ELM, BPNN, RBFNN, and RVM models showed advantages in this study and provided more alternatives for other studies.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This study was supported by National Science and Technology Support Program of China (2014BAD10B02), Zhejiang Provincial Public Welfare Technology Research Projects (2014C32103), Zhejiang Provincial Natural Science Foundation of China (LY15C130003), and the SRF for ROCS, SEM.

References

[1] International Coffee Organization, *World Coffee Trade (1963–2013): A Review of the Markets, Challenges and Opportunities Facing the Sector*, International Coffee Organization, London, UK, 2014.

[2] V. Krivan, P. Barth, and A. F. Morales, "Multielement analysis of green coffee and its possible use for the determination of origin," *Mikrochimica Acta*, vol. 110, no. 4–6, pp. 217–236, 1993.

[3] M. Grembecka, E. Malinowska, and P. Szefer, "Differentiation of market coffee and its infusions in view of their mineral composition," *Science of the Total Environment*, vol. 383, no. 1–3, pp. 59–69, 2007.

[4] A. P. Fernandes, M. C. Santos, S. G. Lemos, M. M. C. Ferreira, A. R. A. Nogueira, and J. A. Nóbrega, "Pattern recognition applied to mineral characterization of Brazilian coffees and sugar-cane spirits," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 60, no. 5, pp. 717–724, 2005.

[5] E. A. De Nadai Fernandes, F. S. Tagliaferro, A. Azevedo-Filho, and P. Bode, "Organic coffee discrimination with INAA and data mining/KDD techniques: new perspectives for coffee trade," *Accreditation and Quality Assurance*, vol. 7, no. 10, pp. 378–387, 2002.

[6] R. M. Alonso-Salces, F. Serra, F. Remero, and K. Heberger, "Botanical and geographical characterization of green coffee (*Coffea arabica* and *Coffea canephora*): chemometric evaluation of phenolic and methylxanthine contents," *Journal of Agricultural and Food Chemistry*, vol. 57, no. 10, pp. 4224–4235, 2009.

[7] A. J. Myles, T. A. Zimmerman, and S. D. Brown, "Transfer of multivariate classification models between laboratory and process near-infrared spectrometers for the discrimination of green Arabica and Robusta coffee beans," *Applied Spectroscopy*, vol. 60, no. 10, pp. 1198–1203, 2006.

[8] I. Esteban-Díez, J. M. González-Sáiz, and C. Pizarro, "An evaluation of orthogonal signal correction methods for the characterisation of arabica and robusta coffee varieties by NIRS," *Analytica Chimica Acta*, vol. 514, no. 1, pp. 57–67, 2004.

[9] I. Esteban-Díez, J. M. González-Sáiz, C. Sáenz-González, and C. Pizarro, "Coffee varietal differentiation based on near infrared spectroscopy," *Talanta*, vol. 71, no. 1, pp. 221–229, 2007.

[10] E. K. Kemsley, S. Ruault, and R. H. Wilson, "Discrimination between *Coffea arabica* and *Coffea canephora* variant robusta beans using infrared spectroscopy," *Food Chemistry*, vol. 54, no. 3, pp. 321–326, 1995.

[11] J. Wang, S. Jun, H. C. Bittenbender, L. Gautz, and Q. X. Li, "Fourier transform infrared spectroscopy for kona coffee authentication," *Journal of Food Science*, vol. 74, no. 5, pp. C385–C391, 2009.

[12] N. Wang, Y. Fu, and L.-T. Lim, "Feasibility study on chemometric discrimination of roasted arabica coffees by solvent extraction and fourier transform infrared spectroscopy," *Journal of Agricultural and Food Chemistry*, vol. 59, no. 7, pp. 3220–3226, 2011.

[13] V. A. Arana, J. Medina, R. Alarcon et al., "Coffee's country of origin determined by NMR: the Colombian case," *Food Chemistry*, vol. 175, pp. 500–506, 2015.

[14] R. Consonni, L. R. Cagliani, and C. Cogliati, "NMR based geographical characterization of roasted coffee," *Talanta*, vol. 88, pp. 420–426, 2012.

[15] Y. B. Monakhova, W. Ruge, T. Kuballa et al., "Rapid approach to identify the presence of Arabica and Robusta species in coffee using ^{1}H NMR spectroscopy," *Food Chemistry*, vol. 182, pp. 178–184, 2015.

[16] R. M. El-Abassy, P. Donfack, and A. Materny, "Discrimination between Arabica and Robusta green coffee using visible micro Raman spectroscopy and chemometric analysis," *Food Chemistry*, vol. 126, no. 3, pp. 1443–1448, 2011.

[17] A. B. Rubayiza and M. Meurens, "Chemical discrimination of arabica and robusta coffees by Fourier transform Raman spectroscopy," *Journal of Agricultural and Food Chemistry*, vol. 53, no. 12, pp. 4654–4659, 2005.

[18] A. Keidel, D. von Stetten, C. Rodrigues, C. Máguas, and P. Hildebrandt, "Discrimination of green arabica and robusta coffee beans by Raman Spectroscopy," *Journal of Agricultural and Food Chemistry*, vol. 58, no. 21, pp. 11187–11192, 2010.

[19] B. Van Eerdenbrugh and L. S. Taylor, "Application of mid-IR spectroscopy for the characterization of pharmaceutical systems," *International Journal of Pharmaceutics*, vol. 417, no. 1-2, pp. 3–16, 2011.

[20] F. Aouidi, N. Dupuy, J. Artaud et al., "Rapid quantitative determination of oleuropein in olive leaves (*Olea europaea*) using mid-infrared spectroscopy combined with chemometric analyses," *Industrial Crops and Products*, vol. 37, no. 1, pp. 292–297, 2012.

[21] A. B. Snyder, C. F. Sweeney, L. E. Rodriguez-Saona, and M. M. Giusti, "Rapid authentication of concord juice concentration in a grape juice blend using Fourier-Transform infrared spectroscopy and chemometric analysis," *Food Chemistry*, vol. 147, pp. 295–301, 2014.

[22] E. Borràs, M. Mestres, L. Aceña et al., "Identification of olive oil sensory defects by multivariate analysis of mid infrared spectra," *Food chemistry*, vol. 187, pp. 197–203, 2015.

[23] S. Hou, C. B. Riley, C. A. Mitchell et al., "Exploration of attenuated total reflectance mid-infrared spectroscopy and multivariate training to measure immunoglobulin G in human sera," *Talanta*, vol. 142, pp. 110–119, 2015.

[24] E. De Luca, S. Bruni, D. Sali, V. Guglielmi, and P. Belloni, "In situ nondestructive identification of natural dyes in ancient textiles by reflection fourier transform mid-infrared (FT-MIR) spectroscopy," *Applied Spectroscopy*, vol. 69, no. 2, pp. 222–229, 2015.

[25] S. Wartewig and R. H. H. Neubert, "Pharmaceutical applications of Mid-IR and Raman spectroscopy," *Advanced Drug Delivery Reviews*, vol. 57, no. 8, pp. 1144–1170, 2005.

[26] A. Edelmann, J. Diewok, K. C. Schuster, and B. Lendl, "Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts," *Journal of Agricultural and Food Chemistry*, vol. 49, no. 3, pp. 1139–1145, 2001.

[27] C. J. Bevin, R. G. Dambergs, A. J. Fergusson, and D. Cozzolino, "Varietal discrimination of Australian wines by means of mid-infrared spectroscopy and multivariate analysis," *Analytica Chimica Acta*, vol. 621, no. 1, pp. 19–23, 2008.

[28] G. Gurdeniz, B. Ozen, and F. Tokatli, "Classification of Turkish olive oils with respect to cultivar, geographic origin and harvest year, using fatty acid profile and mid-IR spectroscopy," *European Food Research and Technology*, vol. 227, no. 4, pp. 1275–1281, 2008.

[29] M. Barker and W. Rayens, "Partial least squares for discrimination," *Journal of Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.

[30] P. Cunningham and S. J. Delany, "k-nearest neighbour classifiers," in *Multiple Classifier Systems*, pp. 1–17, Springer, 2007.

[31] E. K. Kemsley, *Discriminant Analysis and Class Modelling of Spectroscopic Data*, Wiley, Chichester, UK, 1998.

[32] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[33] B. J. Wythoff, "Backpropagation neural networks: a tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 18, no. 2, pp. 115–155, 1993.

[34] L. C. Jain, U. Halici, I. Hayashi, S. B. Lee, and S. Tsutsui, *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, CRC Press, Boca Raton, Fla, USA, 1999.

[35] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, "Extreme learning machine: algorithm, theory and applications," *Artificial Intelligence Review*, vol. 44, no. 1, pp. 103–115, 2013.

[36] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[37] M. N. Murty and V. S. Devi, *Pattern Recognition: An Algorithmic Approach*, Springer Science & Business Media, Dordrecht, The Netherlands, 2011.

[38] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.

[39] W. Dong, Y. Ni, and S. Kokot, "A near-infrared reflectance spectroscopy method for direct analysis of several chemical components and properties of fruit, for example, Chinese hawthorn," *Journal of Agricultural and Food Chemistry*, vol. 61, no. 3, pp. 540–546, 2013.

[40] J. Nogales-Bueno, J. M. Hernández-Hierro, F. J. Rodríguez-Pulido, and F. J. Heredia, "Determination of technological maturity of grapes and total phenolic compounds of grape skins in red and white cultivars during ripening by near infrared hyperspectral image: a preliminary approach," *Food Chemistry*, vol. 152, pp. 586–591, 2014.

[41] O. Galtier, O. Abbas, Y. Le Dréau et al., "Comparison of PLS1-DA, PLS2-DA and SIMCA for classification by origin of crude petroleum oils by MIR and virgin olive oils by NIR for different spectral regions," *Vibrational Spectroscopy*, vol. 55, no. 1, pp. 132–140, 2011.

[42] F. J. Rodríguez-Pulido, D. F. Barbin, D.-W. Sun, B. Gordillo, M. L. González-Miret, and F. J. Heredia, "Grape seed characterization by NIR hyperspectral imaging," *Postharvest Biology and Technology*, vol. 76, pp. 74–82, 2013.

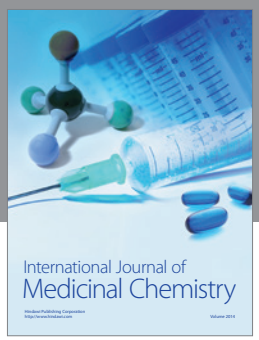
[43] X. Li and Y. He, "Discriminating varieties of tea plant based on Vis/NIR spectral characteristics and using artificial neural networks," *Biosystems Engineering*, vol. 99, no. 3, pp. 313–321, 2008.

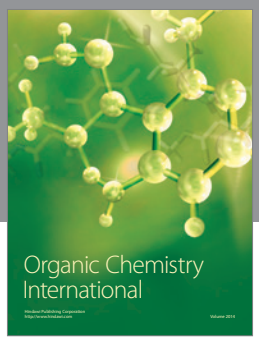
[44] P. Vermeulen, J. A. Fernández Pierna, O. Abbas, P. Dardenne, and V. Baeten, "Origin identification of dried distillers grains with solubles using attenuated total reflection Fourier transform mid-infrared spectroscopy after in situ oil extraction," *Food Chemistry*, vol. 189, pp. 19–26, 2015.

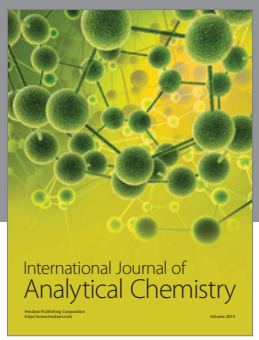
[45] C. Barron, "Prediction of relative tissue proportions in wheat mill streams by fourier transform mid-infrared spectroscopy," *Journal of Agricultural and Food Chemistry*, vol. 59, no. 19, pp. 10442–10447, 2011.

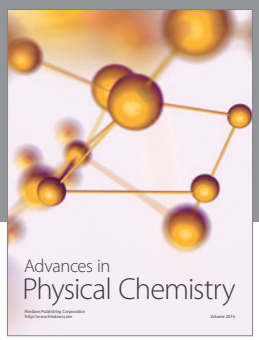
[46] Z. Z. Wu, E. B. Xu, J. Long et al., "Monitoring of fermentation process parameters of Chinese rice wine using attenuated total reflectance mid-infrared spectroscopy," *Food Control*, vol. 50, pp. 405–412, 2015.

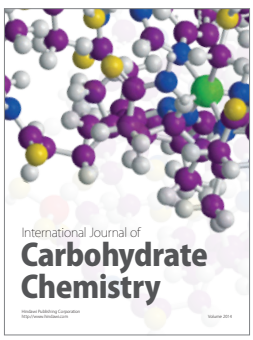
[47] R. M. Balabin, R. Z. Safieva, and E. I. Lomakina, "Gasoline classification using near infrared (NIR) spectroscopy data: comparison of multivariate techniques," *Analytica Chimica Acta*, vol. 671, no. 1-2, pp. 27–35, 2010.

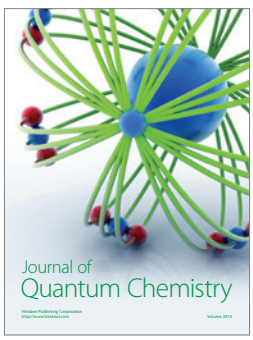


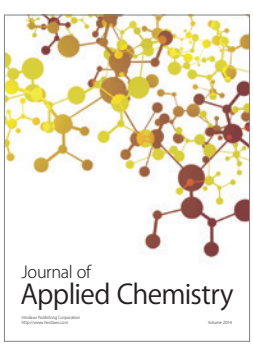


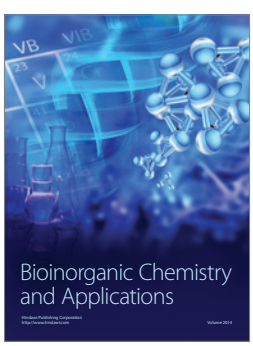


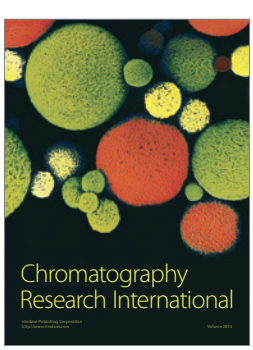


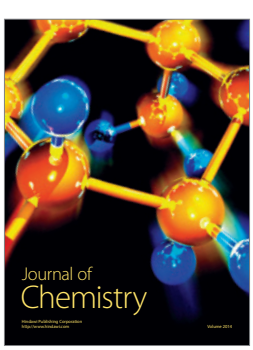












Hindawi

Submit your manuscripts at
http://www.hindawi.com