Identification of True and False Aksu Apple Based on NIRS and PLS-DA

LV Chenfei¹, YANG Jinghui¹*, LIU Yanjun¹, YANG Renjie², ZHANG Chao³ and WU Nan²

¹College of Horticulture and Landscape, Tianjin Agricultural University, Tianjin, 300384, China
²College of Engineering and Technology, Tianjin Agricultural University, Tianjin, 300384, China
³Tianshi College, Tianjin, 301700, China

*Corresponding author’s e-mail: jinghuiyang2@aliyun.com

Abstract: A method for the identification of true and false Aksu Red Fuji apple varieties based on near-infrared spectroscopy (NIRS) and partial least squares discrimination (PLS-DA) was proposed and established. Authentic Xinjiang Aksu Red Fuji and ordinary Red Fuji (Shandong and Shanxi Red Fuji) were purchased from supermarkets. There were 42 samples in each variety, with totaling 126 samples. Near infrared diffuse reflectance spectra of all samples were collected in the range of 4000-12000 cm⁻¹. On the basis of studying the characteristics of NIRS spectra, the principal component analysis (PCA) was carried out. It is pointed out that PCA method cannot fully realize the identification of three kinds Red Fuji Apples. Finally, the partial least squares discriminant analysis model of three kinds Red Fuji Apples was established, and the correct discrimination rates of the three kinds Red Fuji Apples were 100% for samples in calibration set and prediction set.

1. Introduction

Red Fuji apple is the most famous apple variety in the world because of its high quality, delicious taste and storability[1]. Common types of Red Fuji on the market are Aksu Red Fuji in Xinjiang, Yantai Red Fuji in Shandong and Shanxi Red Fuji in Shanxi. The price difference is huge. In order to get more benefits, some businessmen sell ordinary red Fuji as high-value Xinjiang Aksu Red Fuji. Therefore, how to effectively distinguish different kinds of Red Fuji in the market is of great significance to safeguard consumers’ rights and interests.

Near-infrared spectroscopy technology has fast detection speed, no damage, no pollution, and simple operation [2-8], and has become an important means for qualitative identification of agricultural products. Jha[9] used near infrared spectroscopy combined with partial least squares model to identify mango varieties; Kim[10] quickly identified five commercial strawberry varieties based on Fourier transform infrared (FT-IR) data of leaves and fruits and hierarchical tree map of PCA; Mourya[11] established a non-destructive testing model for soybean near infrared reflectance spectroscopy (NIRS) with different genetic backgrounds in 18 countries; Nahm[12] used FT-IR and multivariate analysis to select Silage Maize Varieties; Cortés[13] identified two nectarine cultivars by visible and near infrared diffuse reflectance spectroscopy (VIS-NIR); Xi Wang[14] established a partial least squares regression
(PLSR) model based on near infrared spectroscopy and ultraviolet-visible reference method for rapid identification of green tea samples; He Yong[15] identified snake fruit and other varieties by near infrared spectroscopy based on PCA and neural network; Li Xian Xia[16] uses near-infrared spectroscopy combined with WNN to identify apple species; However, the identification of the true and false Aksu Red Fuji Apple varieties based on NIRS and PLS-DA has not been reported.

2. Test materials and methods

2.1. Test materials

Three different types of test sample (authentic Xinjiang Aksu Red Fuji, Shandong Red Fuji and Shanxi Red Fuji) were purchased from supermarkets, all samples are screened, 42 samples were selected for each variety, totaling 126 samples. Number the filtered samples. Each sample is uniform in size and no obvious damage. It is cleaned and dried and placed in the laboratory environment to be tested for 24 hours.

2.2. Spectrum Acquisition

The test uses a Fourier transform near-infrared spectrometer from PerkinElmer, USA, and the diffuse reflection acquisition accessory is an integrating sphere. The spectrometer was preheated for 15 min. After deducting the background, the sample spectra were collected one by one in the near-infrared band 4000-12000 cm\(^{-1}\) (three times diffuse reflectance spectra were collected from the equatorial portion of the apple, and averaged), and the interval was 2 cm\(^{-1}\). The number of scans is 64 and the spectral resolution is 8 cm\(^{-1}\). A total of 126 red Fuji apple near-infrared spectra were collected (as shown in Figure 1).

2.3. Sample Set Partition

The sample set was divided by KS (Kennard-Stone) method. From the samples of Shandong Red Fuji, Shanxi Red Fuji and Xinjiang Aksu Red Fuji, a representative sample of 84 samples was selected for PLS model establishment and the remaining 42 samples were used as verification sets for classification model validation.

2.4. Establishment and verification of PLS-DA

The experiment used the partial least squares algorithm Matlab code written by the research group to establish the discriminant model of true and false Aksu Red Fuji. Establishment of PLS-DA model: based on the PLS regression method, the regression model is established by using the independent variable matrix X and the classification variable Y of the calibration set samples, and the classification of the samples is judged according to the PLS prediction value of the samples to be classified. PLS-DA model discrimination process: (1) Establish a categorical variable of the calibration set sample (2) PLS analysis of categorical variables and spectral data, establishing a PLS model between spectral data and spectral data (3) The categorical variable value (y\(_{\text{pre}}\)) of the prediction set is
calculated from the PLS model of the categorical variables and spectral features established by the correction set.

The specific criteria are: The Shandong Red Fuji, Shanxi Red Fuji, Xinjiang Aksu Red Fuji classification variables are assigned to -1,0,1; (1) When \(-1.5 < y_{\text{predict}} < 0.5\), it was determined that the samples belonged to the first category, Shandong Red Fuji (2) When \(-0.5 < y_{\text{predict}} < 0.5\), it was determined that the samples belonged to the second category, Shanxi Red Fuji (3) When \(0.5 < y_{\text{predict}} < 1.5\), it was determined that the samples belonged to the third category, Xinjiang Aksu Red Fuji.

2.5. Data Processing
Data processing is carried out by using MATLAB and Unscrambler software. Origin is used for image processing.

3. Analysis and discussion

3.1. Near Infrared Diffuse Reflectance Spectra of True and False Aksu Red Fuji Apple
Figure 2 is a diffuse reflectance spectrum of Xinjiang Aksu Red Fuji, Shandong Red Fuji and Shanxi Red Fuji in the near infrared range of 4000-12000 cm\(^{-1}\). From Figure 2, we can see that the spectral shape and peak position of the three kinds of apple samples are almost the same in the whole spectrum range, so the difference of the three kinds of apple samples cannot be seen by the intuitive spectrum.

![Near Infrared Spectroscopy of True and False Aksu Red Fuji Apple](a) Shandong Red Fuji b) Shanxi Red Fuji c) Xinjiang Red Fuji

Figure 2. Near Infrared Spectroscopy of True and False Aksu Red Fuji Apple

3.2. The principal component analysis of true and false Aksu Red Fuji Apple
The above three types of Red Fuji apples were clustered by principal component analysis (PCA). Figure 3 is a plot of the scatter of the sample of the first three principal components after the principal component decomposition of the original spectral data. From Figure 3, it can be seen that the boundaries of different types of samples are not obvious, there are many overlapping phenomena, and it is impossible to directly distinguish varieties. In order to accurately distinguish the three types of red Fuji, it is necessary to use the supervisory discrimination method to further distinguish.
3.3. Partial Least Square Discriminant Analysis of True and False Aksu Red Fuji Apple

The calibration set samples were subjected to PLS regression and internal cross-validation in the 4000~12000cm\(^{-1}\) spectral range, and the root mean square error (RMSECV) of the interaction verification was used to select the optimal principal component number of the model. Figure 4 is a scatter line diagram of RMSECV varying with the number of principal components (PCs). It can be seen from Figure 4 that when the principal component is 10, the RMSECV of the model is no longer significantly reduced, so the number of principal components of the PLS-DA model is determined to be 10, and a discriminant model is established.

Using the Matlab code of the partial least squares algorithm written by the research group, the 84 calibration set sample data selected by the K-S method were modeled and internally tested. The internal prediction gave an RMSEc of 0.051. From Table 1 and Figure 5, it can be seen that the internal prediction results of the model correction set are correct. All three types of samples have been correctly identified, and the discrimination accuracy is 100%. The model has good self-prediction ability, but the actual prediction ability of the model needs to be further verified by the prediction set.

The 42 samples selected by K-S method were used as prediction sets, and the prediction model of the established classification model was predicted. The predicted RMSEp was 0.131. It can be seen from Figure 6 that the external prediction results of the model prediction set are accurate, and the three types of samples are correctly identified, and the correct rate is 100%. The true and false Aksu Red Fuji apple classification model established by PLS-DA has high prediction accuracy and prediction stability.
4. Conclusion
In this paper, a method for identifying true and false Aksu Red Fuji apple varieties by combining NIRS with PLS-DA is proposed. The experimental results show that the model identification rate established by this method reaches 100%. The method has the advantages of accuracy, simplicity and rapidity, and will certainly provide reference for the non-destructive identification of other fruits and agricultural by-products.

Acknowledgments
We offer our thanks to Tianjin Science and Technology Commission Project (16YFZCNC00750), Largescale breeding and planting project for high quality provenance of oil peony (17ZXBFNC00310), Central Finance Tianjin Major Agricultural Technology Promotion Project (NGT GZD201705), Tianjin Linguo Modern Agricultural Industry Technology System Innovation Team Project (ITTFPRS2018002), Innovative Project for College students (201710061061), Genetic Engineering course Construction Project of Tianjin Agricultural University (2017YKC001), Tianjin Science and Technology Commission Enterprise Specialist Project (18JCTPJJC65300).
References

[1] Mir, J.I., Ahmed, N., Singh, D.B. et al. Diversity evaluation of fruit quality of apple (Malus × domestica Borkh.) germplasm through cluster and principal component analysis. Ind J Plant Physiol. (2017) 22: 221.

[2] Lee, JD., Shannon, J.G. & Choung, MG. Selection for protein content in soybean from single F2 seed by near infrared reflectance spectroscopy. Euphytica (2010) 172: 117.

[3] Song, H.J., Kim, Y.D., Jeong, M.J. et al. Rapid selection of theanine-rich green tea (Camellia sinensis L.) trees and metabolites profiling by Fourier transform near-infrared (FT-IR) spectroscopy. Plant Biotechnol Rep (2015) 9: 55.

[4] Joshi, A., Raigond, P. & Singh, B. Near-infrared spectroscopy: a non-destructive tool for processing quality evaluation of potatoes. Ind J Plant Physiol. (2017) 22: 258.

[5] Rodríguez-Zúñiga, U.F., Farinas, C.S., Carneiro, R.L. et al. Fast Determination of the Composition of Pretreated Sugarcane Bagasse Using Near-Infrared Spectroscopy. Bioenerg. Res. (2014) 7: 1441.

[6] Khan, K.M., Krishna, H., Majumder, S.K. et al. Detection of Urea Adulteration in Milk Using Near-Infrared Raman Spectroscopy. Food Anal. Methods (2015) 8: 93.

[7] Pereira, J.M., Leme, L.M., Perdoncini, M.R.F.G. et al. Fast Discrimination of Milk Contaminated with Salmonella sp. Via Near-Infrared Spectroscopy. Food Anal. Methods (2018) 11: 1878.

[8] Jha, S.N. & Garg, R. Non-destructive prediction of quality of intact apple using near infrared spectroscopy. J Food Sci Technol (2010) 47: 207.

[9] Jha, S.N., Jaiswal, P., Narsaiah, K. et al. Authentication of Mango Varieties Using Near-Infrared Spectroscopy. Agric Res (2013) 2: 229.

[10] Kim, S.W., Min, S.R., Kim, J. et al. Rapid discrimination of commercial strawberry cultivars using Fourier transform infrared spectroscopy data combined by multivariate analysis. Plant Biotechnol Rep (2009) 3: 87.

[11] Mourya, V., Kumar, V., Rani, A. et al. Near-Infrared Reflectance Spectroscopy for Protein Content in Soybean Flour and Screening of Germplasm Across Different Countries. Agric Res (2016) 5: 29.

[12] Nahm, S.H., Yang, S.G., Kim, S.W. et al. Rapid Discrimination of F1 Hybrid Seeds from Their Parental Lines and Selection of Protein-Rich Corn Lines for Silage Corn Breeding Using FT-IR Spectroscopy Combined by Multivariate Analysis. J. Crop Sci. Biotechnol. (2015) 18: 161

[13] Cortés, V., Blasco, J., Aleixos, N. et al. Visible and Near-Infrared Diffuse Reflectance Spectroscopy for Fast Qualitative and Quantitative Assessment of Nectarine Quality. Food Bioprocess Technol (2017) 10: 1755.

[14] X. Wang, J. Huang, W. Fan et al. Identification of green tea varieties and fast quantification of total polyphenols by near-infrared spectroscopy and ultraviolet-visible spectroscopy with chemometric algorithms. Anal. Methods (2015) 7: 787-792

[15] HE Yong, LI Xiao-li, SHAO Yong-ni. Discrimination of Varieties of Apple Using Near Infrared Spectra Based on Principal Component Analysis and Artificial Neural Network Model [J]. Spectroscopy and Spectral Analysis, 2006(05):850-853.

[16] LI Xian-Xia, CHEN Fei, ZHANG Jing-jing. Identification of Apple Types by Infrared Spectroscopy with Wavelet Neural Networks [J]. Chinese Journal of Spectroscopy Laboratory, 2012, 29(06):3529-3531.