ACS OMEGA

http://pubs.acs.org/journal/acsodf

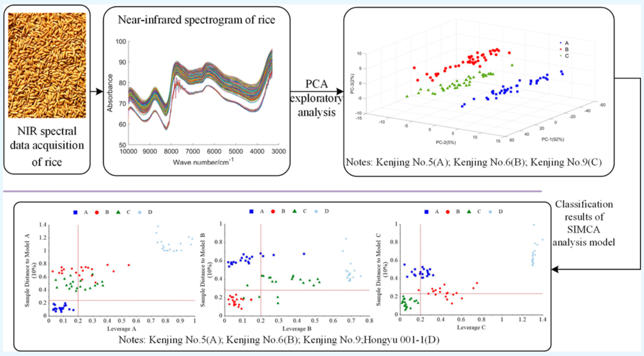
Article

Classification of Rice Varieties Using SIMCA Applied to NIR Spectroscopic Data

GuoJun Shi, XiaoWen Zhang, Ge Qu, and ZhengGuang Chen*

Cite This: *ACS Omega* 2022, 7, 46623−46628

ABSTRACT: The selection of suitable rice varieties is the key to achieve high and stable yields, and the correct identification of rice varieties is the prerequisite for seed selection. In this paper, with Kenjing No.5, No.6, and No.9 as the subjects, the effectiveness of near-infrared spectroscopy (NIRS) combined with soft independent modeling of class analogy (SIMCA) in the rapid identification of rice varieties was explored. The modeling sets of Kenjing No.5, No.6, and No.9 samples were respectively used to establish a SIMCA classification model based on principal component analysis (PCA). The accuracies of the model in classifying the rice samples in the modeling set were 100, 100, and 97.5%, respectively. Then, the established SIMCA model was used to identify the rice samples in the test set. According to the experimental findings, the SIMCA analytical method achieved 100% prediction accuracy for the Kenjing No.5, Kenjing No.6, and Hongyu 001−1 samples. For the Kenjing No.9 sample, the accuracy rate was 90% with a 10% sample of Kenjing No.9 misidentified as Kenjing No.6. Therefore, the analytical method of NIRS combined with SIMCA could effectively identify the rice varieties, providing a new approach for the correct selection of planting varieties.



1. INTRODUCTION

Rice is one of the grain crops that human beings live on. In China alone, there have been more than 50,000 varieties of rice, a number that keeps rising with the progress in research on new hybrid rice varieties.[1] As the selection of suitable rice varieties is one of the keys that determine the economic benefits and breeding of rice, the rapid identification of rice varieties turns out to be an essential topic in today's agricultural production, crop breeding, and seed testing.[2] Primarily affected by the climate and geographical circumstances, the rice varieties do not vary greatly within the same region and cannot be identified simply through the rice seed's appearance and morphology, resulting in tremendous financial costs. At present, the commonly used identification methods for the rice varieties around the world include morphology, fluorescence scanning identification,[3] chemical identification,[4] and electrophoresis identification. Among them, the morphological approach has limited application and low accuracy. Although the remaining three feature higher accuracy, they are highly demanding on the operator's specific skills and take a lot of time for processing. As a result, none of these methods apply to bulk analysis and nondestructive online detection of the samples.[5] The four rice varieties chosen in this study (Kenjing No.5, Kenjing No.6, Kenjing No.9, and Hongyu 001−1) are commonly cultivated in Heilongjiang Province, China, and each variety is suited for growing in a distinct cumulative temperature zone. Because there are minimal changes in appearance between the four kinds, it is impossible to tell them apart with the naked eye. As a result, a quick, nondestructive, and low-cost approach for classifying rice types is required.

In recent years, near-infrared spectroscopy (NIRS) has been widely applied in many industries including agriculture,[6] food,[7,8] pharmaceuticals,[9,10] and industry.[11,12] SIMCA (Soft Independent Modeling of Class Analog) is a traditional pattern recognition model that primarily uses the class distance between PCA models of various classes to ascribe the class of samples. The SIMCA model is able to give good results in the case of whether the sample belongs to a certain class or not.[13,14] In recent years, NIR spectroscopy combined with SIMCA has been widely used in sample classification.[14] Suhandy[15] used a near-infrared spectral analysis technique combined with the SIMCA model to classify six specialty coffees from Indonesia based on variety and geographical origin. PCA was employed for unsupervised exploratory

Received: August 29, 2022
Accepted: November 30, 2022
Published: December 9, 2022

https://doi.org/10.1021/acsomega.2c05561
ACS Omega 2022, 7, 46623−46628

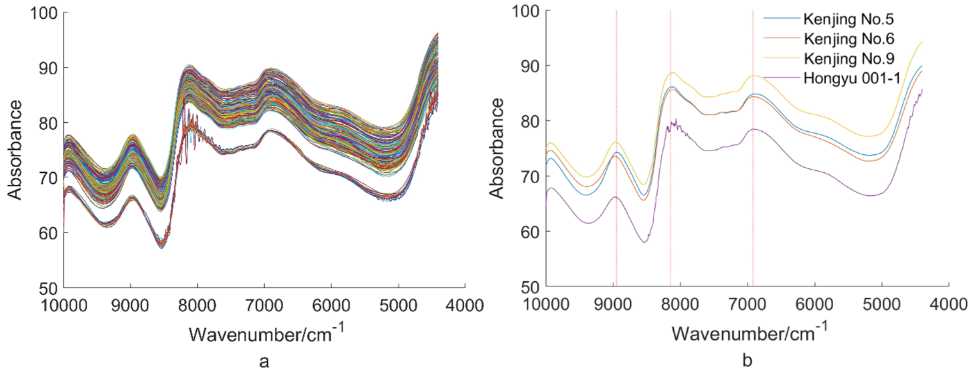


Figure 1. Near-infrared spectrum curve (a) and average spectrum curve (b) of different rice varieties.

analysis[16] and soft independent modeling of class analogy (SIMCA) for supervised classification. The results demonstrated that the SIMCA model correctly classified most of the tested sample sets into the corresponding categories. Khodabakhshian[17] used an analytical method relying on near-infrared spectroscopy (NIR) in combination with SIMCA model to classify the two main impurities in turmeric powder (Sudan red and metanil yellow) and established SIMCA model. The correct classification rates of samples in training set and test set were 96.7 and 93.4, respectively. It is proved that the SIMCA model is effective to classify the data acquired by near-infrared spectroscopy. Shirzadifar et al.[18] used NIR spectroscopy combined with the SIMCA model to classify three common weeds with 100% identification accuracy. The results showed that the NIR spectroscopy combined with the SIMCA model had a good discriminatory ability for the three common weeds. The above study demonstrates the feasibility of the NIR spectroscopic analysis technique combined with the SIMCA approach in sample classification.

Currently, rice varieties are mainly identified by manual or gas chromatography[19] analytical techniques, both of which have issues with difficult operation, arduous processing, and poor detection accuracy. As a quick, environmentally friendly, and nondestructive analytical technique, NIR spectroscopy analytical technique can quickly and easily gather vital data on experimental samples that contain hydrogen, such as transmittance, absorbance, reflectance, and chemical bond strength. This information is crucial for the identification of rice varieties using NIR. Therefore, an analytical technique based on NIRS that does not require pretreatment was proposed in combination with the SIMCA model as a reliable analytical method[20] to identify the rice varieties more rapidly.

2. MATERIALS AND METHODS

2.1. Materials and Equipment. All 200 samples from the four rice types, including Kenjing No.5, Kenjing No.6, Kenjing No.9, and Hongyu 001−1, which were harvested in the autumn of 2016 at the experimental base of Heilongjiang Bayi Agricultural University (Location: 46.59°N, 125.16E), were used in this study. Following sampling, it was discovered that there were 60, 60, 60, and 20 of each of the four types, respectively. The rice was planted on May 15, 2016, and harvested on October 30, 2016. The WQF-600N-type Fourier transform NIR spectrometer (Beijing Rayleigh Analytical Instrument Co. Ltd, China) was adopted, with a wave range of 10,000−3300 cm^{-1}, a resolution of 8 cm^{-1}, and a scanning number of 16 times. The Unscrambler X 10.3 (CAMO Software AS, NORWAY) was employed for spectrum analysis.

2.2. NIR Measurement and Sample Set Division. The rice samples of three varieties were put alongside with the NIR spectrometer in the same indoor environment for 24 hours, to ensure consistent environmental conditions for both the rice and the instrument. After preheating the spectrometer for 10 min, the instrument's background was first collected. Then, the sample was spread on the bottom of the glass containers to collect the spectral data. Each sample was scanned 16 times to obtain the mean value. Spectral data were collected by measurement of diffuse reflectance from the homogenized sample in the NIR region within 10,000−4000 cm^{-1}. Figure 1a shows the NIR spectra of the four rice varieties, and Figure 1b indicates the average spectral curves of the four rice varieties. From the figure, it can be seen that the trends of the changes in the spectral reflectance of the rice seed samples of different varieties are the same, but the reflectance differs, with the lower reflectance of the rice seeds of the Hongyu 001−1 variety and the higher reflectance of the rice seeds of the Kenjing No.9 variety. There are three distinct characteristic peaks at 8950, 8150, and 6920 cm^{-1} in rice seeds. The absorption peak at 8950 cm^{-1} is mainly generated by the C−H bond second harmonic generation and O−H bond second harmonic generation. The absorption peak at 8150 cm^{-1} is mainly the second-order frequency doubling of the C−H bond. The absorption peak at 6920 cm^{-1} is mainly the first-order frequency doubling absorption peak of O−H bond stretching vibration.[21] Figure 1a demonstrates that the spectra of Hongyu 001−1 sample differ significantly from those of the other three varieties due to the high level of noise in the region 7700−8200 cm^{-1}. This is a result of the diversity, which includes variations in sample appearance and chemical makeup.

In this study, two-thirds of the samples were randomly selected from the sample sets of Kenjing No.5, No.6, and No.9 (i.e., 40 samples randomly selected for each variety) to establish the modeling set and 60 samples for test (about 120 samples altogether). As the test set, we used the remaining third of the sample sets Kenjing No.5, Kenjing No.6, Kenjing No.9, and Hongyu 001−1 (about 120 samples altogether). As the test set, we used the remaining third of the sample sets Kenjing No.5, Kenjing No.6, Kenjing No.9, and Hongyu 001−1(each group has 20 samples, for a total of 80 samples.).

2.3. Soft Independent Modeling of Class Analogue. SIMCA classification, also known as similarity and analogy model, is a supervised pattern recognition model based on a known class model.[22] Basically, it processes to create a PCA model for each class in the modeling set, thereby reducing the

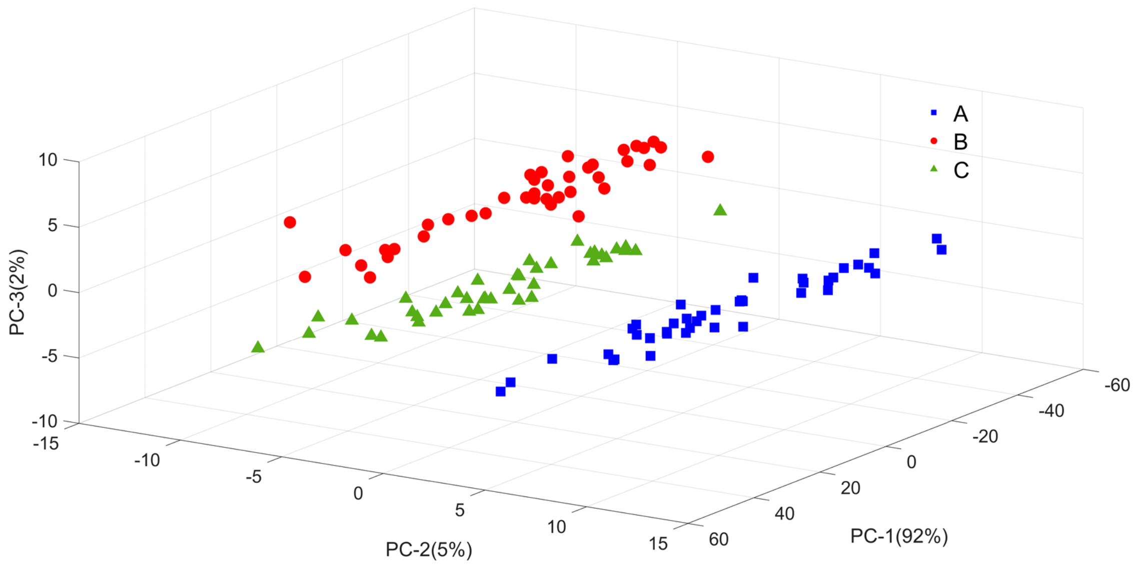


Figure 2. PCA score plots of modeling sample sets.

dimensions of the data. The aim of PCA is to obtain the score matrix T and loading matrix P of each rice variety in a known class. After that, the distance between the unknown sample and each centroid class in the model is calculated, and the distance identification method is used to identify the class of unknown samples. In chemometrics, the distance between samples is usually used to represent the difference between samples. The basic idea of discriminant analysis is to compare the class distances from a sample to centroid each class, and the class with the shortest distance is exactly the samples.[23] The calculation commonly uses two kinds of distances: Euclidean distance and mahalanobis distance.[24] The Euclidean distance is employed in distance identification by SIMCA classification.[25] The calculation can be divided into two steps.[21,26]

2.4.1. Step One. Principal component analysis was conducted on the training set, and various SIMCA principal component analysis models were established (Formula 1)

$$\boldsymbol{X}_k = \boldsymbol{T}_k\boldsymbol{P}_k^t + \boldsymbol{E}_k \tag{1}$$

where $\boldsymbol{X}_k$ is the spectral matrix $(n \times m)$, n is the number of all samples of category k, m is the number of wavelength variables, $\boldsymbol{T}_k$ is the scoring matrix $(n \times f)$, f is the best principal component, $\boldsymbol{P}_k$ is the load matrix $(m \times f)$, and $\boldsymbol{E}_k$ is the spectral residual matrix $(n \times m)$.

2.5.1. Step Two. For the test sample set (new samples of unknown types), a left cross-test method is used to validate the SIMCA model established by the training set. Calculate the orthogonal distance (OD (Formula 2)) from the new sample y_{new} to the center of the class model (class k model). The model with the lowest OD value is the model that y_{new} belongs to. To achieve the classification of new samples

$$\text{OD} = \| y_{\text{new}} - \hat{y} \| \tag{2}$$

where $\hat{y}$ denotes the projection of y_{new} onto the PCA model.

$$\hat{y} = \bar{x} + \text{PP}'(y_{\text{new}} - \bar{x}) \tag{3}$$

where $\bar{x}$ denotes the mean spectrum of the modeling sample set.

3. RESULTS AND DISCUSSION

3.1. PCA Analysis. Before modeling, PCA was carried out on 120 samples (40 samples for each variety) of modeling set, and the score plot of the modeling sample set is shown in Figure 2. The first three principal components explain 97% of the variance in the original data. In the score plots of PC1 and PC2, some sample points are overlapped (Figure 2), but in the score plots of PC2 and PC3, the sample points show a clear trend of clustering (Figure 2). Therefore, the modeling set can be used to establish a classification model for rice varieties on the basis of SIMCA. There is a significant difference between the three varieties. In particular, Kenjing No.5(A) is relatively far from the other two, while Kenjing No.6(B) and No.9(C) are close to each other.

PCA models were established for the three varieties in the modeling set individually respectively. Each of these PCA models contained the corresponding variety's feature information, and there was no outlier among the samples, with a good degree of fitting. SIMCA prediction model was established on the basis of choosing 3 as the optimal number of components for the three models. The distances between the three known sample sets of the prediction model are shown in Table 1. As suggested by the table, all of the distances between different sample sets were greater than 3, which could also be reflected in Figure 2. In particular, the model distance between Kenjing No.6 and No.9 was the smallest, i.e., 4.55 (larger than 3). Therefore, there were significant differences between each two models, that is, they could be easily distinguished from each other.

Table 1. Distances between Models

sample	Kenjing No.5	Kenjing No.6	Kenjing No.9
Kenjing No.5(A)	1	19.42	13.23
Kenjing No.6(B)	19.42	1	4.55
Kenjing No.9(C)	13.23	4.55	1

3.2. Supervised Classification of SIMCA. SIMCA classification is a binary classification model, which only results in yes or no. In other words, a sample either belongs to a class or not. Since three different varieties were used in this

https://doi.org/10.1021/acsomega.2c05561
ACS Omega 2022, 7, 46623−46628

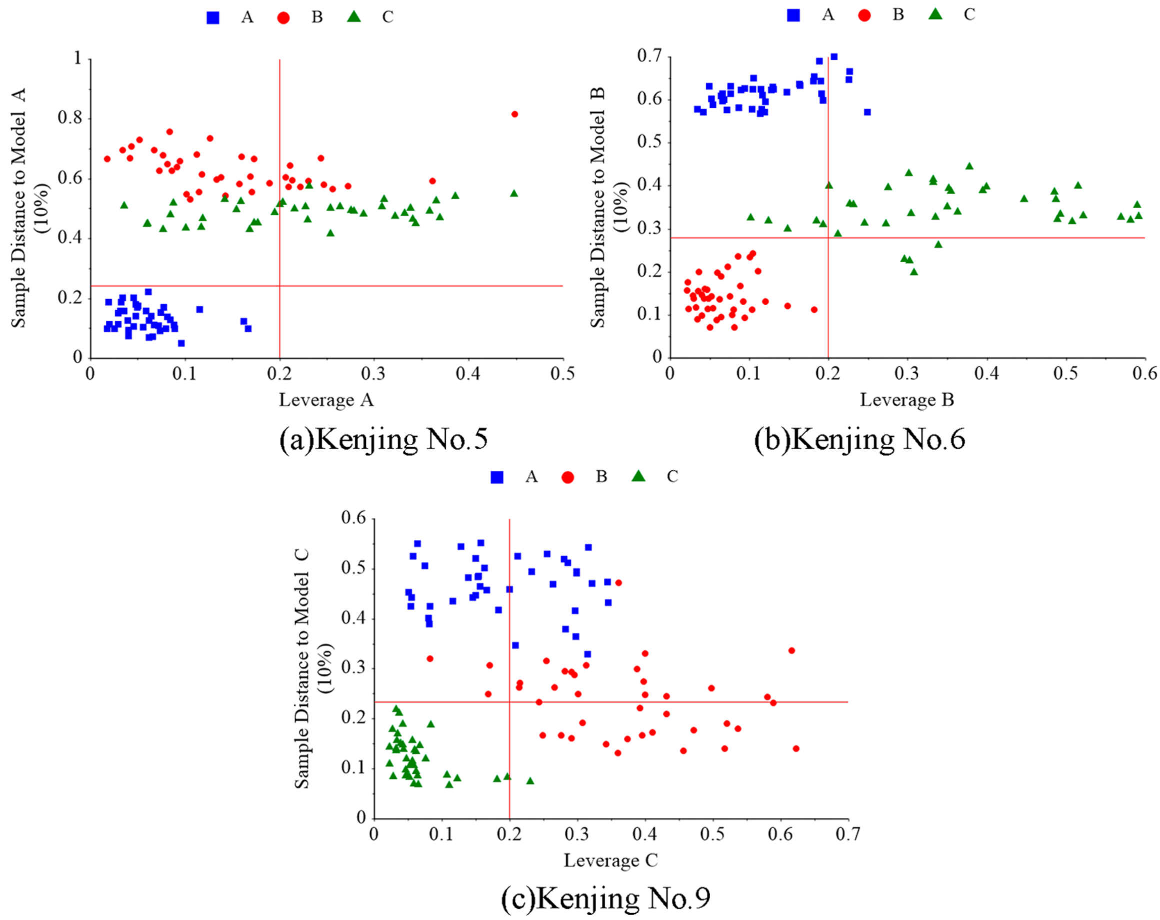


Figure 3. Identified results of three varieties in the modeling set.

study, the class of each sample should be separately identified. In this study, through identification of the modeling sample set, the accuracy of SIMCA classification model was validated. Figure 3 shows the identified results of three varieties in the modeling set by the SIMCA classification model. In the figure, the horizontal axis represents the leverage value of the sample, that is, the distance from the sample to the model center. It shows the difference between the sample and other sample points in the class, without considering the degree of description for the sample by the class model. The vertical axis represents the distance from the sample to the model, which is calculated using the square root of the sample residuals. The smaller the value, the more accurate the model.[21] As suggested by Figure 3a,b, SIMCA could provide a good identification result. At a significance level of α = 10%, all modeling sample sets of Kenjing No.5(A) and No.6(B) were correctly identified, and all samples fell within the quadrilateral area defined by the two mutually perpendicular lines and the axes. Figure 3c shows the SIMCA identification result of Kenjing No.9(C). One of the samples had a great leverage value that was far beyond the quadrilateral area. However, since the sample was far from the other two models, it was not classified into the other varieties. Instead, the sample was identified as not belonging to any variety.

All modeling sample sets were used to validate the accuracy of the SIMCA classification model, with the statistical results shown in Table 2. The accuracy of SIMCA model in identifying Kenjing No.5, No.6, and No.9 was 100, 100, and 97.5% at the 10% significance level, respectively. It means that the misjudgment rate for Kenjing No.9 was 2.5%. This was because some samples were omitted, possibly due to the fact that the modeling sample set contained insufficient information and thus led to insignificant difference between the two classes of samples.

Table 2. Accuracy of SIMCA Classification Model

modeling sample set (n = 120)	Kenjing No.5/%	Kenjing No.6/%	Kenjing No.9/%
identification accuracy (α = 10%)	100	100	97.5

3.3. SIMCA Model Predicting Verification Set Samples. SIMCA model was used to classify the 80 samples in the test set. To validate the effectiveness of the model, among the 80 samples, there were another 20 samples that did not belong to any of the existing classes in addition to the 60 calibration samples of Kenjing No.5(A), No.6(B), and No.9(C) (20 samples of each). The variety of the added samples was actually Hongyu 001−1(D), which was not previously involved in SIMCA modeling, Figure 4 shows the prediction results of the test set samples in four classes at the 10% significance level, where the diamonds represent the Hongyu 001−1 samples.

As suggested by Figure 4a,b, all of the Kenjing No.5 and No.6 among the test set samples were correctly classified. However, in Figure 4b, two Kenjing No.9 samples were mistakenly wrongly divided into Kenjing No.6, and the varieties could not be completely separated. As suggested by Figure 4a−c, all of the Hongyu 001−1 samples among the unknown rice samples were not mistakenly classified into any

https://doi.org/10.1021/acsomega.2c05561
ACS Omega 2022, 7, 46623−46628

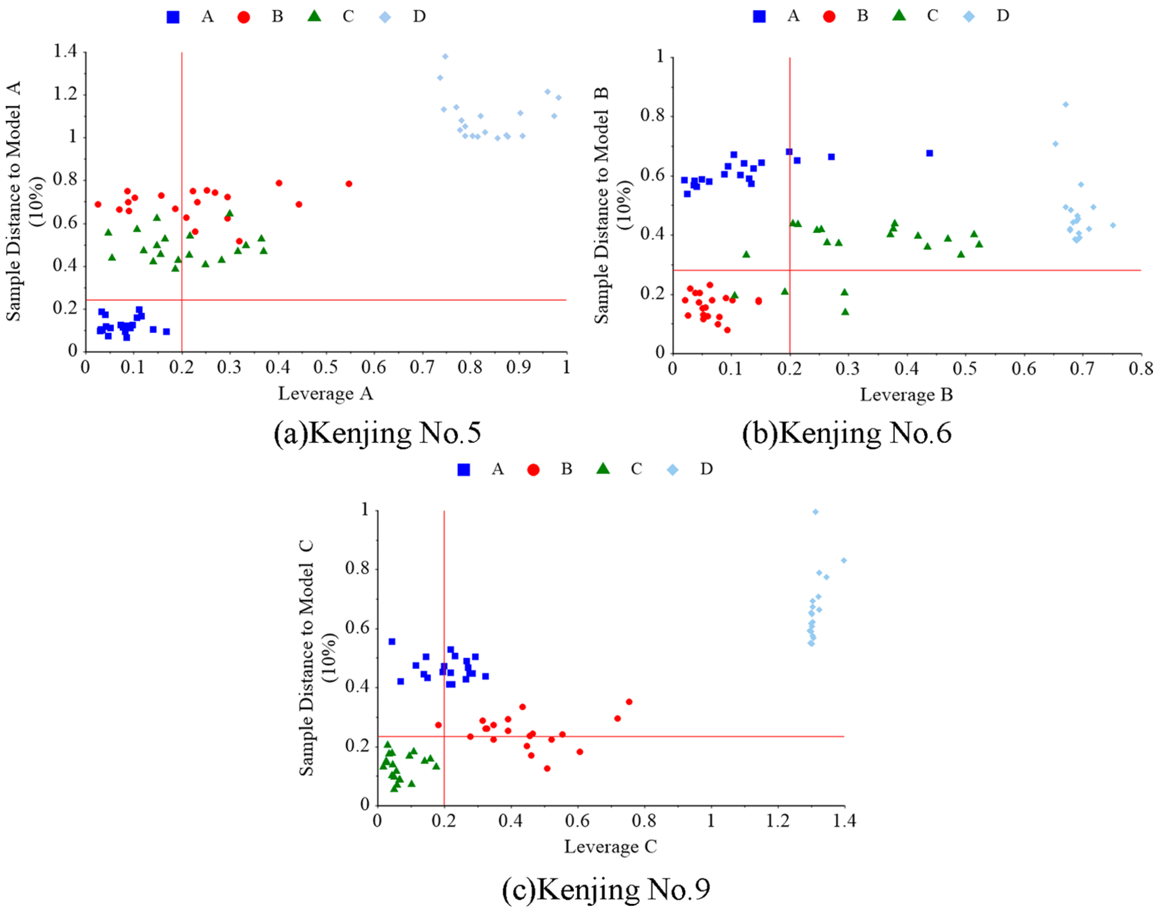


Figure 4. Verifies the sample classification results.

known class, which in turn indicated the effectiveness of the proposed model.

The accuracy of the SIMCA model in the prediction of unknown samples is shown in Table 3. The prediction accuracy of the SIMCA model for Kenjing No.5, No.6, and No.9 was 100, 100, and 90% at the 10% significance level, respectively. Namely, there was only a misjudgment rate of 10% for Kenjing No.9. Given that none of the Hongyu 001−1 samples that had been added in the test set as new unknown variety was mistakenly identified as any one of the Kenjing No.5, No.6, and No.9 varieties, it could be implied that the misjudgment of Kenjing No.6 and No.9 was caused by the insignificant difference in chemical properties between the two varieties, which further proved the effectiveness of the proposed method.

Table 3. Accuracy of SIMCA Model in Prediction of the Test Set Samples

test set sample (n = 60)	Kenjing No.5(A)/%	Kenjing No.6(B)/%	Kenjing No.9(C)/%
Identification accuracy (α = 10%)	100	100	90

According to the above results, we can see that the SIMCA model can be used to identify rice varieties. After the misjudged samples were removed from the original SIMCA model, the classification model was rebuilt, but the accuracy in classifying unknown samples was not significantly improved. The main reasons for this may be that the information covered by the sample was insufficiently thorough, and Kenjing 9 seed was cultivated on the basis of Kenjing 6 seed, with only minor chemical property variations between the seeds.

4. CONCLUSIONS

The combination of NIRS and SIMCA could effectively distinguish between the rice varieties with little difference in appearance and components. Experimental results showed that the accuracy of the SIMCA classification model in identifying Kenjing No.5, No.6, and No.9 in the modeling sample set was 100, 100, and 97.5%, respectively. For the prediction of unknown rice samples (including Kenjing No.5, No.6, No.9, and Hongyu 001−1), the model's accuracy for Kenjing No.5, No.6, and No.9 was 100, 100, and 90%, respectively. Specifically, all of the Hongyu 001−1 among the unknown rice sample were not mistakenly identified as any one of Kenjing No.5, No.6, and No.9 varieties. Finally, it was speculated that the primary cause for misjudgment might be that the samples did not contain sufficient information and that the difference in chemical properties between Kenjing No.6 and No.9 was insignificant. By increasing the information contained by each set of samples or by multiplying the number of modeling sample sets, the PCA model of each variety could be made more representative so as to improve the accuracy of the SIMCA classification model. The findings showed that the combination of NIRS and SIMCA could rapidly and reliably identify the rice varieties, and thus the proposed analytical method could be implemented in and guide the agricultural production of rice.

https://doi.org/10.1021/acsomega.2c05561
ACS Omega 2022, 7, 46623−46628

AUTHOR INFORMATION

Corresponding Author

ZhengGuang Chen − *College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China;* orcid.org/0000-0002-6791-4240; Email: ruzee@byau.edu.cn

Authors

GuoJun Shi − *College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China*

XiaoWen Zhang − *College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China*

Ge Qu − *College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c05561

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was supported by the Scientific Research Foundation for Returned Scholars, Heilongjiang Bayi Agricultural University (XDB201813). The authors thank the Heilongjiang Bayi Agricultural University for financial support while undertaking this research project.

REFERENCES

(1) Muazam, A.; Praptana, R. H. Seeds Phenotypes Characterization of Promising Tungro-Resistant Rice Lines. *IOP Conf. Ser. Earth Environ. Sci.* **2022**, *985*, 012054.

(2) Luo, Z.; Linying, Z.; Tianbing, C.; Muhua, L.; Jinyin, C.; Huamao, Z.; Mingyin, Y. Rapid identification of rice species by laser-induced breakdown spectroscopy combined with pattern recognition. *Appl. Opt.* **2019**, *58*, 1631.

(3) Zhang, Y.; Jinan, X.; Chaomin, Z.; Ming, L.; Feifei, C. Characterization of the Stability of Vegetable Oil by Synchronous Fluorescence Spectroscopy and Differential Scanning Calorimetry (DSC). *Anal. Lett.* **2021**, *54*, 2376−2686.

(4) Chen, J.; Li, M.; Pan, T.; Pang, L.; Yao, L.; Zhang, J. Rapid and non-destructive analysis for the identification of multi-grain rice seeds with near-infrared spectroscopy. *Spectrochim. Acta, Part A* **2019**, *219*, 179−185.

(5) Liu, S.; Chen, Z.; Jiao, F. Rapid identification of the variety of maize seeds based on near-infrared spectroscopy coupled with locally linear embedding. *Appl. Opt.* **2022**, *61*, 1704−1710.

(6) Cui, Y.; Lanjun, X.; Dong, A.; Zhe, L.; Jiancheng, G.; Shaoming, L.; Xiaodong, Z.; Dehai, Z. Identification of maize seed varieties based on near infrared reflectance spectroscopy and chemometrics. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 177−183.

(7) Sai, X.; Jinchang, R.; Huazhong, L.; Xu, W.; Xiuxiu, S.; Xin, L. Nondestructive detection and grading of flesh translucency in pineapples with visible and near-infrared spectroscopy. *Postharvest Biol. Technol.* **2022**, *192*, 112029.

(8) Silva, L. C. R.; Folli, G. S.; Santos, L. P.; Barros, I. H. A. S.; Oliveira, B. G.; Borghi, F. T.; dos Santos, F. D.; Filgueiras, P. R.; Rom, W. Quantification of beef, pork, and chicken in ground meat using a portable NIR spectrometer. *Vib. Spectrosc.* **2020**, *111*, 103158.

(9) Kranenburg, R. F.; Yannick, W.; Martin, A.; Sharon, S.; Peter, W. F. A.; Annette, E.; Peter, H. J. K.; Arian, C. v. A. The importance of wavelength selection in on-scene identification of drugs of abuse with portable near-infrared spectroscopy. *Forensic Chem.* **2022**, *30*, 100437.

(10) Lee, C.; Juan, J. P.; Kirsten, E. K.; Selina, C. W. Near-Infrared (NIR) Spectrometry as a Fast and Reliable Tool for Fat and Moisture Analyses in Olives. *ACS Omega* **2018**, *3*, 16081.

(11) Gonçalves, M.; Ribeiro, M.; Paiva, N. T.; Ferra, J. M.; Martins, J.; Magalhães, F.; Carvalho, L. Resin and volatile content of melamine-impregnated paper assessed by near infrared spectroscopy, a simulation of the industrial process using a laboratory-scale gantry. *Eur. J. Wood Wood Prod.* **2020**, *78*, 1279−1286.

(12) Valinger, D.; Tamara, J.; Adela, Š.; Nikolina, M.; Manuela, P.; Maja, B.; Jasenka, G. K.; Ana, J. T. Development of ANN models based on combined UV−vis−NIR spectra for rapid quantification of physical and chemical properties of industrial hemp extracts. *Phytochem Anal.* **2020**, *32*, 326−338.

(13) Rácz, A.; Gere, A.; Bajusz, D.; Héberger, K. Is soft independent modeling of class analogies a reasonable choice for supervised pattern recognition? *RSC Adv.* **2018**, *8*, 10−21.

(14) Alexey, L. P.; Oxana, Y. R. New Trends in Qualitative Analysis: Performance, Optimization, and Validation of Multi-Class and Soft Models. *TrAC, Trends Anal. Chem.* **2021**, 116372.

(15) Suhandy, D.; Yulia, M. Authentication of Six Indonesian Ground Roasted Specialty Coffees According to Variety and Geographical Origin using NIR Spectroscopy with Integrating Sphere. *IOP Conf Ser Earth Environ Sci.* **2021**, *830*, 012065.

(16) Mazivila, S. J.; Olivieri, A. C. Chemometrics coupled to vibrational spectroscopy and spectroscopic imaging for the analysis of solid-phase pharmaceutical products: A brief review on non-destructive analytical methods. *Trends Anal. Chem.* **2018**, *108*, 74−87.

(17) Khodabakhshian, R.; Reza, B. M.; Bagher, E. Adulteration detection of Sudan Red and metanil yellow in turmeric powder by NIR spectroscopy and chemometrics: The role of preprocessing methods in analysis. *Vib. Spectrosc.* **2022**, *120*, 103372.

(18) Shirzadifar, A.; Bajwa, S.; Mireei, S. A.; Howatt, K.; Nowatzki, J. Weed species discrimination based on SIMCA analysis of plant canopy spectral data. *Biosyst. Eng.* **2018**, *171*, 143−154.

(19) Ju, X.; Feiyu, L.; Hongyi, G.; Yuying, J.; Yuan, Z.; Degang, X. Identification of Rice Varieties and Adulteration Using Gas Chromatography-Ion Mobility Spectrometry. *IEEE Access* **2021**, *9*, 18222.

(20) Escandar, G. M.; Goicoechea, H. C.; Peña, A. M. d. l.; Olivieri, A. C. Second- and higher-order data generation and calibration: A tutorial. *Anal. Chim. Acta* **2014**, *806*, 8−26.

(21) XiaoLi, Z.*Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Applications*; Chemical Industry Press: BeiJing, 2011.

(22) Panero, F. d. S.; Oscar, S.; João, S. P.; Fernando, S. D. V. F.; dos, S. P. P.; Anselmo, F. R. R. Non-Destructive Genotyping of Cultivars and Strains of Sesame through NIR Spectroscopy and Chemometrics. *Biosensors* **2022**, *12*, 69.

(23) Firmani, P.; Bucci, R.; Marini, F.; Biancolillo, A. Authentication of "Avola almonds" by near infrared (NIR) spectroscopy and chemometrics. *J. Food Compos. Anal.* **2019**, *82*, 103235.

(24) Das, S.; Baruah, H. K. A Comparison of Two Fuzzy Clustering Techniques. *J. Diabetes Sci Technol.* **2013**, *1*, 1−15.

(25) Wu, P.; Song, H.; Yang, W.; Zhang, Z.; Han, X. The Origin of Eggs Based on Near Infrared Spectroscopy. *Shipin Gongye Keji* **2020**, *41*, 227−231.

(26) Képeš, B. E.*Classification of Metals by Means of Laser-induced Breakdown Spectroscopy and Chemometric Methods*, Brno University of Technology, 2016/17.

https://doi.org/10.1021/acsomega.2c05561
ACS Omega 2022, 7, 46623−46628