Honey quality detection based on near-infrared spectroscopy

Man ZHOU[1], Tao LONG[2], Zhengyang ZHAO[2], Jie CHEN[2] (iD), Qingsong WU[2], Yue WANG[2], Zhiyong ZOU[2*]

Abstract

As a natural agricultural product, honey is favored by consumers, and its variety and adulteration have a huge impact on the quality. Acacia honey, red jujube honey and rape honey were used as experimental objects, and their spectral reflectance curves were obtained through a near-infrared spectral image acquisition system. Spectral features were extracted from the preprocessed spectral reflectance curves, and a honey variety classification model based on near-infrared spectral features was established by machine learning. After statistical analysis, Principal Component Analysis Support Vector Machine after processing data through Successive Projections Algorithm (SPA-SVM) is the optimal classification model for three varieties of acacia honey, red jujube honey and rape honey, and the correct rate of honey variety classification reaches 95.83%. The spectral reflectance curve was used to establish a honey adulteration identification model based on the partial least squares-discriiminate analysis (PLS-DA), and the classification accuracy was 97.92% in the test set.

Keywords: honey; quality; machine learning; adulteration.

Practical Application: Honey detection by near infrared spectroscopy.

1 Introduction

In recent years, consumers pay more attention to the quality and nutrition of honey (Rasad et al., 2018). The variety and authenticity of honey directly affect the quality of honey (Phillips & Abdulla, 2022). In the current honey market, there is a phenomenon of shoddy and fake. For profit, some unscrupulous merchants randomly label honey varieties, confusing the concept of varieties with high nutritional value and inferior varieties of honey. Or incorporate low-cost vegetable syrups in the sales process. It not only has a negative impact on the establishment of honey brands, but also damages the rights and interests of consumers (Naila et al., 2018).

There is an increasing demand for alternative analytical techniques in the search for fast, accurate and reliable quality control systems for mass production to prevent fraud and thus ensure food safety (Cagri-Mehmetoglu, 2018). Most food and agricultural products research based on HSI adopts hyperspectral reflectance imaging technology (Hossen et al., 2021), which measures the reflectance from visible light area to short wave infrared area (Pingzhen et al., 2022). Many studies have been applied in honey adulteration. Phillips et al., 2022 established the adulteration detection model by using the new honey of hyperspectral imaging and machine learning, and the classification accuracy reached more than 95% (Phillips & Abdulla, 2022). Zhang & Abdulla (2022) used hyperspectral imaging to establish a new Zealand honey plant source classification model, with classification accuracy of more than 99%, and established different spectral characteristics analysis (Zhang & Abdulla, 2022). Shao et al. (2022) established hyperspectral imaging for nondestructive detection of honey adulteration. The LIBSVM model has a classification accuracy of 92.5% for honey adulteration, realizing the detection of honey adulteration. A partial least squares regression (PLSR) model was used to establish a prediction model

for the level of adulteration. The verification accuracy of the model was 0.84, and the root mean square error (RMSEV) of the verification was 5.26% (Shao et al., 2022).

Hyperspectral imaging technology has been applied in many fields, such as peanut (Zou et al., 2022b) germination prediction (Zou et al., 2022c), mildew detection (Zou et al., 2022a), hot pot (Zou et al., 2023) oil detection (Zou et al., 2022d), fruit grading (Zou et al., 2021), etc. Using physical and chemical methods to judge the quality of honey (Kek et al., 2017; Wan et al., 2018), the results are often very accurate (Shamsudin et al., 2019). However, the cost is high, the time-consuming process (Jimenez et al., 2016) is long, and the original structure of the honey needs to be destroyed. It is often not a good method for varietal and adulteration discrimination on large batches of honey. As one of the most rapidly developing high-tech analytical techniques in recent years, near-infrared spectroscopy (Lei et al., 2021; Miao et al., 2021) has been widely used in the quality inspection of agricultural products due to its rapidity (Wang et al., 2021; Yang et al., 2021), non-destructiveness and simplicity (Wu et al., 2016). Taking a variety of honeys sold on the market in Ya'an, Sichuan as the research object, the research on the identification of honey varieties and the adulteration of single honey varieties with different mass fractions of fructose syrup by near-infrared spectroscopy was discussed.

2 Materials and methods

2.1 Sample acquisition and preparation

The trial consisted of two phases, and the honey samples used in the first phase were all honey from 2017. Including 3 different varieties of acacia honey, red jujube honey and rape honey in

Ya'an City, Sichuan Province, the total number of samples is 120, including 40 acacia honey, 40 red jujube honey, and 40 acacia honey. As shown in Table 1.

The honey used in the second stage is the honey produced in 2018. When preparing adulterated honey, the collected liquid rape honey samples were 20% and 40% by mass, and the high fructose syrup of the variety Yihai Kerry F60 was added to the normal rape honey samples according to different mass ratios. 40 adulterated rape honey samples containing 20% fructose syrup by mass fraction and 40 adulterated rape honey samples containing 40% sugar syrup were obtained. Table 2 shows the quality components of pure rape honey and blended high fructose syrup contained in the three types of samples.

Accurately weigh 10 g (± 0.05 g) samples of each variety with a high-precision electronic scale. Use a 60 mm high transparent thickened petri dish to evenly hold the honey sample and evenly cover the bottom of the petri dish. 40 samples were collected from each class for model building and validation. Before spectrum collection, the samples with crystals were first treated with a 50 °C water bath to dissolve the crystals in them, and then placed at room temperature overnight to collect spectrum data (room temperature: 17 ± 5 °C).

2.2 Hyperspectral image acquisition

The experiment uses the ImSpector series hyperspectrometer of Zhuoli Hanguang Company, whose spectral effective band range is 387-1035 nm, the band resolution is 2.8 nm, and there are a total of 256 bands. The region of interest in the corrected spectral image of the oil sample was intercepted by ENVI5.1 (Exelis Visual Information Solutions Inc., USA) software. Then calculate the average value of the pixels on the region of interest in the image corresponding to each wavelength band from 387 to 1035 nm as the characteristic reflectance spectrum curve of the honey sample.

2.3 Black and white correction

When the near-infrared spectral imaging system is in different wavelength bands, the noise intensity of honey samples is different. When the light source intensity is weak, the noise is more strongly affected. In addition, the spectral curve

acquisition is also affected by the dark current in the camera. Therefore, performing black and white correction is the first step in acquiring spectral images.

2.4 Hyperspectral data preprocessing

In order to eliminate the influence of environmental factors and non-quality factor information of the spectral equipment itself in the collected hyperspectral data, and obtain more valuable spectral data, it is necessary to preprocess the black and white corrected spectral data (Jha & Garg, 2010). Five spectral preprocessing methods were used to improve the signal-to-noise ratio. They are Mean centering, Autoscaling, Normalization, Standard Normalized Variate (SNV), Multiple Scattering Correction.

2.5 Data dimensionality reduction

The near-infrared spectral imaging system can collect spectral images at 256 consecutive wavelength sampling points in the wavelength range of 380-1040 nm of a sample. At the same time, due to the large sample size in the experiment, the data volume of the entire spectral matrix is particularly large. Such a complex data is mixed with a lot of repeated data information, and even contains noise data that adversely affects the model results. To eliminate these useless data and redundant data that affect the speed of model building and the accuracy of results, the idea is to reduce the dimensionality of the data.

Principal component analysis

Principal Component Analysis (PCA) (Hao, 2021) is a widely used data analysis tool that reduces the dimensionality of a set of variables and still preserves as much of the original information as possible (Wu & Chyu, 2004). PCA transforms the variables into a new coordinate system called the principal component space. The maximum variance of the data projection in the principal component space is on the first coordinate axis (referred to as the first principal component PC1 at this point). The coordinate where the second largest variance of the data is located is called the second principal component (PC2). In the new coordinate space, the principal components (or principal factors) are orthogonal to each other and have the property of being uncorrelated. After transformation, the overlapping information in the original data is eliminated (Richards, 1988). The first principal component PC1 indicates that when the variance of the original data matrix is the largest, it contains the most information.

Successive projections algorithm

The Successive Projections Algorithm (SPA) (Pontes et al., 2005; Soares et al., 2013) is one such method based on feature

Table 1. Source and number of honey samples.

Variety	Origin	Number
Acacia honey	Ya'an, Sichuan Province	40
red jujube honey	Yulin, Shaanxi Province	40
rape honey	Ya'an, Sichuan Province	40

Table 2. Quality in the sample.

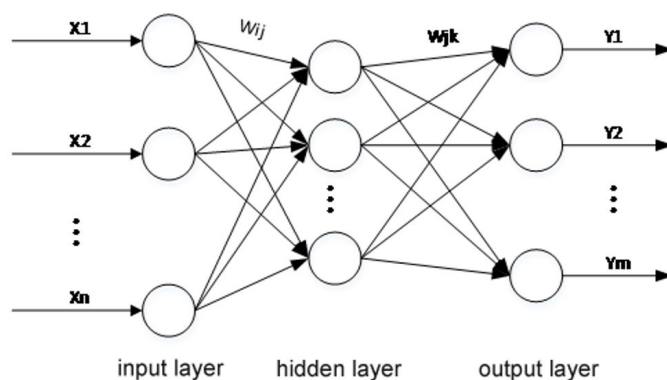| sample type	Contains pure canola honey quality	Incorporated high fructose syrup quality
Pure Rape Honey	10.00 g	0.00 g
20% Fructose Syrup with Adulterated Canola Honey	8.00 g	2.00 g
40% Fructose Syrup with Adulterated Canola Honey	6.00 g	4.00 g

Figure 1. BP neural network topology diagram.

extraction, which can be used to deal with classification problems. The theoretical basis of the SPA algorithm is to find the feature group with the least redundant information in the entire original data space through the observation and calculation of the original variable. Finally, the multicollinearity between the features in the feature group is minimized. SPA is simpler and less time-consuming in the wavelength selection process (Shi et al., 2014).

2.6 Model building

60% of each type of spectral curve data of 120 honey samples was randomly selected as the training set, and the remaining 40% of the spectral curve data was used as the test set. Machine learning is used to build a classification model, and support vector machine (SVM) and BP neural network algorithms are used to build model analysis. The confusion matrix was used to evaluate the stability and reliability of the model.

Based on the confusion matrix, we further introduce four True Positive Rate (TPR), False Negative Rate (FNR), False Positive Rate (FPR), and True Negative Rate (TNR) index. At the same time, in order to reduce various random errors in the test process, unless otherwise specified, all the results about the accuracy in the paper are the results obtained after running 5 times and averaging.

Support vector machine algorithm

Support Vector Machine (SVM) was first proposed by Corinna Cortes and Vladimir Vapnik in 1995. It is used to establish a supervised learning model for data classification and regression analysis (Sudershan & Rao, 2020), and has strong adaptability and generalization ability (Kashef, 2021). SVM (Chen & Lin, 2008) adopts the method of finding the least structured risk to improve the generalization ability of the model distribution and achieve the purpose of minimizing the empirical risk and confidence range. To achieve the purpose of obtaining good statistical laws from a small amount of data, and still obtain good statistical laws. The SVM model has been shown to be highly scalable and robust (Lu & Wang, 2005).

BP neural network algorithm

BP neural network is the earliest, most widely used and most successful neural network because of its simple structure and the
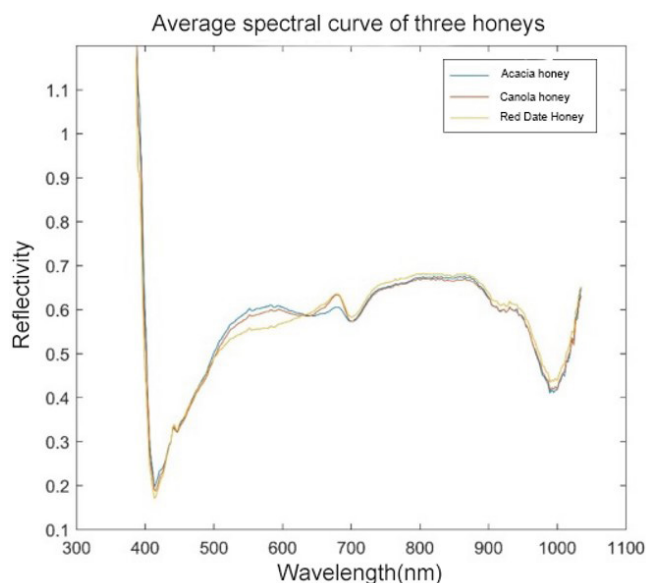

Figure 2. Average spectral curve of three honeys.

ability to deal with complex nonlinear problems (Lyu & Zhang, 2019). BP neural network is a one-way multilayer feedforward network (Song et al., 2021). As shown in Figure 1, the output of the previous layer is used as the input of the next layer, and the error is propagated by the back-propagation algorithm. That is, the gradient of the weight is determined by the error, and the gradient is further updated. The learning process consists of forward propagation of signals and back propagation of errors (Zhang et al., 2018).

3 Results and discussion

3.1 Classification of honey varieties

Spectral analysis of samples

The average spectral curves of the three varieties of honey are shown in Figure 2. The three types of honey have a strong absorption on the spectrum between 400-500 and 600-750 nm, while the absorption on the spectrum is lower at less than 400 nm and greater than 800 nm. The absorption curves of the three types of honeys at 690 nm are obviously different, which may be

related to the sugar content in the three types of honeys. In the whole wavelength range, the wavelength point with the highest degree of difference in the absorption rate of the three types of honey can be selected for data analysis, so as to discriminate the honey varieties.

From Figure 3, the average spectral curves of the three honeys are analyzed, and there are multiple absorption peaks on the entire spectral curve from 380 nm to 1040 nm, and the range covers the entire spectrum. The five absorption peaks at the wavelengths of 413.5 nm, 644.4 nm, 702.5 nm, 914.9 nm and 989.3 nm are the most obvious. And the average spectral curve of three different honeys of acacia honey, red jujube honey and rape honey can be known. Due to the difference in the composition and concentration of organic matter contained in different varieties of honey, the reflectance is different on the spectral curve.

SPA feature wavelength screening

As shown in Figure 4. Represents the change in RMSE value as the number of variables increases, and the boxes represent the number of wavelengths picked out. It can be seen that at the beginning of the selection of sampling points, the root mean square error value has a rapid decline process. When the number of wavelength points is selected as 16, the root mean square error reaches a minimum stable value.

The distribution of the selected 19 characteristic wavelengths over the entire spectral wavelength range is shown in Figure 5. These 19 wavelengths express the range, peak, reflectivity and other related information of the entire spectral curve very well.

The wavelength point data selected by the SPA feature is used as the input of the BP model. The accuracy of the classification results is shown in Table 3 and Table 4.

In the training set, three honey samples of acacia, red dates, and rape are in the model, 21 of 24 acacia honeys are correctly classified, and the remaining 3 are classified as rape honey. The group predictions of red date honey and rape honey were all correct. The classification accuracy of SPA-BP for training samples is only 95.83%.

It shows that the continuous projection feature selection method combined with BP neural network has the best classification effect of 100% on acacia honey. However, it is not accurate enough to predict the effect of jujube honey and rape honey.

The wavelength point data selected by the SPA feature is used as the input of the SVM model. The accuracy of the classification results is shown in Table 5 and Table 6.

In Table 5, the classification accuracy rate of three kinds of honey samples of acacia, red dates and rape by Successive Projections Algorithm combined with Support Vector Machine model (SPA-SVM) for training samples is 100%.

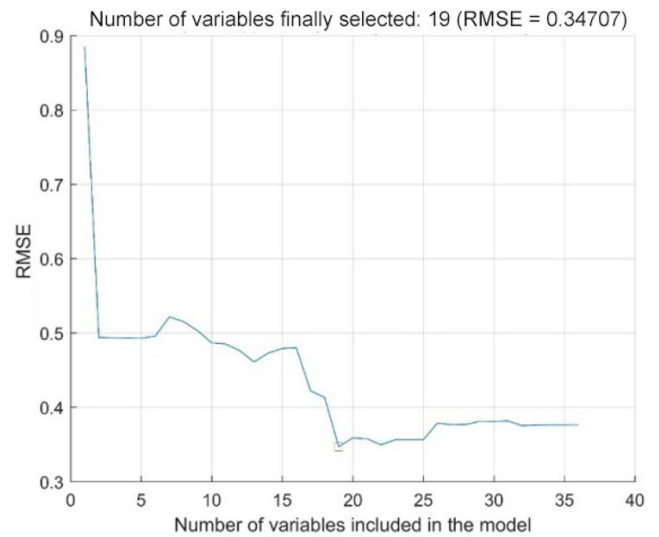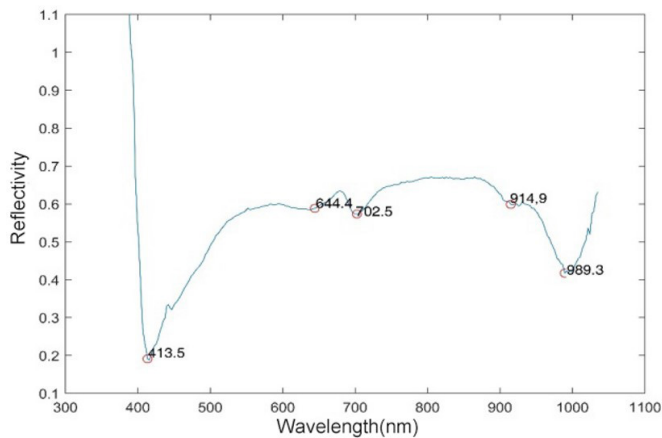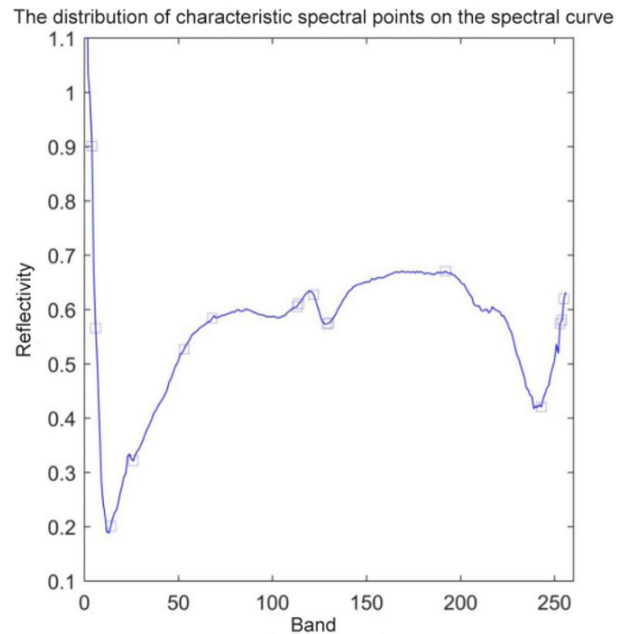**Figure 4**. SPA picks the final number of selected variables.

Figure 5. The distribution of characteristic wavelength points filtered by SPA on the original spectral curve.

Figure 3. Wavelength point of the characteristic absorption peak in the spectrum.

Table 3. Training set analysis results of SPA-BP discriminant model.

Expected	Training set			Discrimination rate	
	Acacia honey	red jujube honey	rape honey	Group discrimination rate	Overall discrimination rate
Acacia honey (n = 24)	21	0	3	87.5%	
red jujube honey (n = 24)	0	24	0	100%	95.83%
rape honey (n = 24)	0	0	24	100%	

SPA-BP: Successive Projections Algorithm - Back Propagation.

Table 4. Test set analysis results of SPA-BP discriminant model.

Expected	Training set			Discrimination rate	
	Acacia honey	red jujube honey	rape honey	Group discrimination rate	Overall discrimination rate
Acacia honey (n = 16)	14	2	0	91.67%	
red jujube honey (n = 16)	4	12	0	83.33%	87.5%
rape honey (n = 16)	0	0	16	100%	

Table 5. Training set analysis results of SPA-SVM discriminant model.

Expected	Training set			Discrimination rate	
	Acacia honey	red jujube honey	rape honey	Group discrimination rate	Overall discrimination rate
Acacia honey (n = 24)	24	0	0	100%	
red jujube honey (n = 24)	0	24	0	100%	100%
rape honey (n = 24)	0	0	24	100%	

Table 6. Test set analysis results of the SPA-SVM discriminant model.

Expected	Training set			Discrimination rate	
	Acacia honey	red jujube honey	rape honey	Group discrimination rate	Overall discrimination rate
Acacia honey (n = 16)	16	0	0	100%	
red jujube honey (n = 16)	1	14	1	87.5%	95.83%
rape honey (n = 16)	0	0	16	100%	

Table 7. Training set analysis results of PCA-BP discriminant model.

Expected	Training set			Discrimination rate	
	Acacia honey	red jujube honey	rape honey	Group discrimination rate	Overall discrimination rate
Acacia honey (n = 24)	24	0	0	100%	
red jujube honey (n = 24)	1	23	0	95.83%	98.61%
rape honey (n = 24)	0	0	24	100%	

PCA-BP: Principal Component Analysis - Back Propagation.

As shown in Table 6, in the test set, 16 acacia honeys were all correctly classified. Among the 16 jujube honeys, 1 was predicted as acacia honey, and 1 was predicted as rape honey. The 16 rape honeys were classified correctly. It shows that SPA-SVM model has the best classification effect of acacia honey and rape honey, and the classification accuracy of red jujube honey is only 87.5%.

PCA feature wavelength screening

Through the spectral curve dimensionality reduction calculation, the cumulative contribution rate of the principal components is shown in Figure 6.

The PCA principal component dimensionality reduction data is used as the input of the BP model, and the classification accuracy is shown in Table 7 and Table 8.

In the PCA-BP model, all samples of acacia and rape honey were completely identified correctly. One of the jujube honey samples was misclassified, misclassifying jujube honey as acacia honey. The classification accuracy rate was 95.83%, and the total discrimination rate was 98.61%.

As shown in Table 8, the classification effect of PCA-BP neural network on rape honey is completely correct, and the effect on red jujube honey and acacia honey is average. The overall discrimination rate of the model is not improved compared to

the BP neural network classification model established by the original spectral data. On the contrary, the extracted principal component factors cannot have no optimization effect on the BP neural network model.

The PCA principal component dimensionality reduction data is used as the input of the SVM model, and the classification accuracy is shown in Table 9 and Table 10.

Through Tables 9-10: the test set in the PCA-SVM discriminant model prediction results: rape honey classification is completely correct. Of the 16 acacia honeys, 2 were predicted to be jujube

honey. Of the 16 red dates, 2 were predicted to be acacia honey. It shows that the classification effect of the PCA-SVM model on rape honey is completely correct, but the classification of acacia honey and red jujube honey is always blurred. The overall discrimination rate of the model is improved to 91.67% compared to the 89.58% classification rate of the support vector machine classification model established by the original spectral data. This shows that the data is reduced by principal component analysis and imported into the least squares support vector machine to optimize the model and improve the classification accuracy of the model.

The final results are shown in Table 11. The accuracy of model discrimination using SVM algorithm is generally higher than that after classification using BP algorithm. The model with the highest accuracy in predicting the classification effect of the three honey products is the SVM model established by using the spectral feature variables filtered by SPA as input, and the classification accuracy is as high as 95.83%. The use of near-infrared spectral features to classify honey varieties is of practical significance.

3.2 Identification of adulteration of honey

Spectral analysis of samples

The analysis steps are similar to those for the classification of honey varieties. Firstly, the average spectral curve of pure honey, 20% high fructose syrup adulterated honey and 40% high fructose syrup adulterated honey was observed. It can be seen from Figure 7 above that the spectral shapes of the three types of samples, as well as the peak and trough trends of spectral absorption, are consistent. However, in the whole spectrum
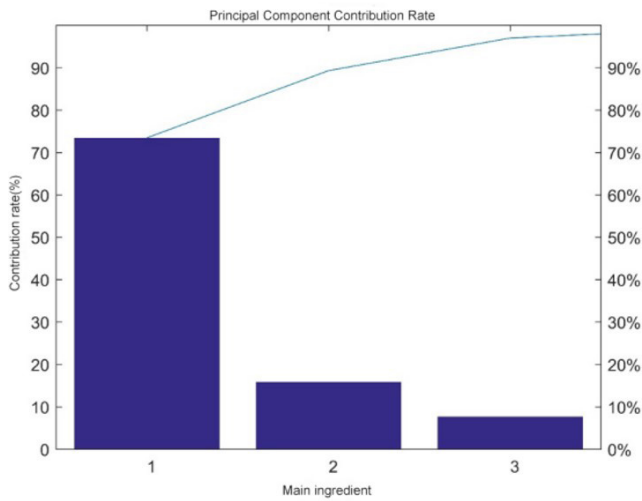

Figure 6. Cumulative contribution rate of the three principal components.

Table 8. Test set analysis results of PCA-BP discriminant model.

Expected	Training set			Discrimination rate	
	Acacia honey	red jujube honey	rape honey	Group discrimination rate	Overall discrimination rate
Acacia honey (n = 16)	14	2	0	87.5%	
red jujube honey (n = 16)	3	12	1	75%	87.5%
rape honey (n = 16)	0	0	16	100%	

Table 9. Training set analysis results of PCA-SVM discriminant model.

Expected	Training set			Discrimination rate	
	Acacia honey	red jujube honey	rape honey	Group discrimination rate	Overall discrimination rate
Acacia honey (n = 24)	24	0	0	100%	
red jujube honey (n = 24)	1	23	0	95.83%	98.61%
rape honey (n = 24)	0	0	24	100%	

Table 10. Test set analysis results of PCA-SVM discriminant model.

Expected	Training set			Discrimination rate	
	Acacia honey	red jujube honey	rape honey	Group discrimination rate	Overall discrimination rate
Acacia honey (n = 16)	14	2	0	87.5%	
red jujube honey (n = 16)	2	14	0	87.5%	91.67%
rape honey (n = 16)	0	0	16	100%	

**Table 11**. Overall classification accuracy of different data processing methods in classification models.

	BP Classification		SVM Classification	
	Training set	test set	Training set	test set
original spectrum	100%	89.58%	100%	89.58%
SPA	95.83%	87.5%	100%	95.83%
PCA	98.61%	87.5%	98.61%	91.67%

Table 12. Training set analysis results of discriminant models established by PLS-DA.

	forecast group			Discrimination rate	
Expected	Pure Rape Honey	20% Fructose Syrup Adulterated Honey	40% Fructose Syrup Adulterated Honey	Group discrimination rate	Overall Discrimination Rate
Pure Rape Honey (n = 24)	24	0	0	100%	
20% Fructose Syrup Adulterated Honey (n = 24)	0	24	0	100%	
40% Fructose Syrup Adulterated Honey (n = 24)	0	0	24	100%	100%

Table 13. Test set analysis results of PLS-DA discriminant model.

	forecast group			Discrimination rate	
Expected	Pure Rape Honey	20% Fructose Syrup Adulterated Honey	40% Fructose Syrup Adulterated Honey	Group discrimination rate	Overall Discrimination Rate
Pure Rape Honey (n = 16)	16	0	0	100%	
20% Fructose Syrup Adulterated Honey (n = 16)	0	16	0	100%	
40% Fructose Syrup Adulterated Honey (n = 16)	0	1	15	93.75%	97.92%
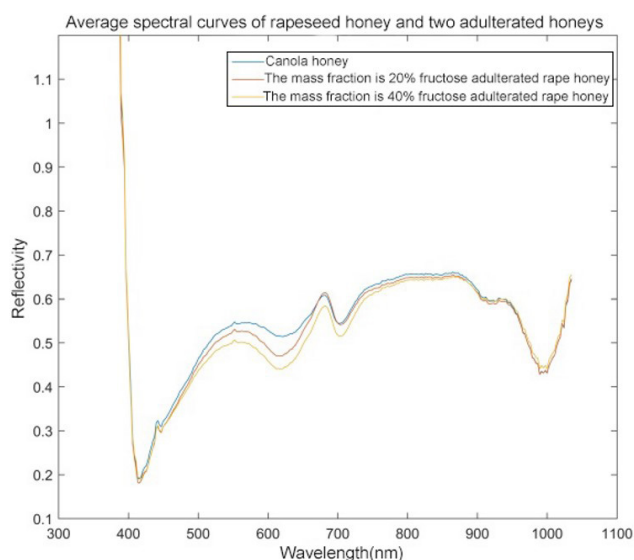
Figure 7. Average spectral curve of pure rape honey and two different adulterated concentrations of rape.

sampling range, there are differences in the reflection values of the three types of samples at the sensitive wavelength point. The light absorption of the three types of samples is mainly concentrated in the range of 500-750 nm, and there are three obvious absorption peaks near 615 nm and 714 nm. At the absorption peak near 615 nm, the reflectance of the three types of samples has a large difference. At the absorption peak near 714 nm, pure rape honey and 20% fructose adulterated honey basically overlap, and 40% adulterated honey is different from the previous two. From the overall average spectral curve, the higher the quality of the adulterated fructose syrup, the stronger the absorption of the spectrum. The performance of the model on the training set is shown in Table 12.

Use the pre-set 48 samples as the test set to make predictions on the established classification model. Combining with Table 13, there is a 40% high fructose syrup adulterated rape honey which is mistakenly divided into 20% high fructose syrup adulterated rape honey. The rest of the validation set samples were correctly classified, and the model prediction accuracy rate was 97.92% (47/48).

4 Conclusion

In order to realize the rapid identification of honey quality while considering various indicators. The near-infrared spectroscopy technology was used to classify and identify the different varieties of honey collected, and to detect the adulteration of pure honey and adulterated honey samples. In the identification of acacia honey, rape honey, and red jujube honey, SPA and PCA were used for data dimensionality reduction. The dimensionality-reduced data was combined with BP and SVM classifiers to

select the most suitable model for the classification of the three honey varieties. The SPA-SVM model has a classification accuracy rate of 95.83%, which is the best classification model in the classification of honey samples. Using PLS-DA to study the discriminant model of whether to participate in honey, the optimal number of main factors was selected. The classification effect of the model classification accuracy rate of 97.92% is achieved. It provides an effective discrimination model between pure honey and different quality fractions of participating honey.

References

Cagri-Mehmetoglu, A. (2018). Food safety challenges associated with traditional foods of Turkey. *Food Science and Technology*, 38(1), 1-12. http://dx.doi.org/10.1590/1678-457x.36916.

Chen, Y.-W., & Lin, C.-J. (2008). Combining SVMs with various feature selection strategies. In I. Guyon, M. Nikravesh, S. Gunn & L. A. Zadeh (Eds.), *Feature extraction: foundations and applications* (pp. 315-324, Studies in Fuzziness & Soft Computing, no. 207). Berlin: Springer.

Hao, W. (2021). Classification of sport actions using principal component analysis and random forest based on three-dimensional data. *Displays*, 72, 102135.

Hossen, M. T., Ferdaus, M. J., Hasan, M. M., Lina, N. N., Das, A. K., Barman, S. K., Paul, D. K., & Roy, R. K. (2021). Food safety knowledge, attitudes and practices of street food vendors in Jashore region, Bangladesh. *Food Science and Technology*, 41(Suppl. 1), 226-239. http://dx.doi.org/10.1590/fst.13320.

Jha, S. N., & Garg, R. (2010). Non-destructive prediction of quality of intact apple using near infrared spectroscopy. *Journal of Food Science and Technology*, 47(2), 207-213. http://dx.doi.org/10.1007/s13197-010-0033-1. PMid:23572626.

Jimenez, M., Beristain, C. I., Azuara, E., Mendoza, M. R., & Pascual, L. A. (2016). Physicochemical and antioxidant properties of honey from Scaptotrigona mexicana bee. *Journal of Apicultural Research*, 55(2), 151-160. http://dx.doi.org/10.1080/00218839.2016.1205294.

Kashef, R. (2021). A boosted SVM classifier trained by incremental learning and decremental unlearning approach. *Expert Systems with Applications*, 167, 114154. http://dx.doi.org/10.1016/j.eswa.2020.114154.

Kek, S. P., Chin, N. L., Yusof, Y. A., Tan, S. W., & Chua, L. S. (2017). Classification of entomological origin of honey based on its physicochemical and antioxidant properties. *International Journal of Food Properties*, 20(sup3), S2723-S2738. http://dx.doi.org/10.1080/10942912.2017.1359185.

Lei, L., Ke, C., Xiao, K., Qu, L., Lin, X., Zhan, X., Tu, J., Xu, K., & Liu, Y. (2021). Identification of different bran-fried Atractylodis Rhizoma and prediction of atractylodin content based on multivariate data mining combined with intelligent color recognition and near-infrared spectroscopy. *Spectrochimica Acta. Part A: Molecular and Biomolecular Spectroscopy*, 262, 120119. http://dx.doi.org/10.1016/j.saa.2021.120119. PMid:34243140.

Lu, W. Z., & Wang, W. J. (2005). Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends. *Chemosphere*, 59(5), 693-701. http://dx.doi.org/10.1016/j.chemosphere.2004.10.032. PMid:15792667.

Lyu, J., & Zhang, J. (2019). BP neural network prediction model for suicide attempt among Chinese rural residents. *Journal of Affective Disorders*, 246, 465-473. http://dx.doi.org/10.1016/j.jad.2018.12.111. PMid:30599370.

Miao, X. X., Miao, Y., Tao, S. H., Liu, D. B., Chen, Z. W., Wang, J. M., Huang, W. D., & Yu, Y. Y. (2021). Classification of rice based on storage time by using near infrared spectroscopy and chemometric methods. *Microchemical Journal*, 171, 106841. http://dx.doi.org/10.1016/j.microc.2021.106841.

Naila, A., Flint, S. H., Sulaiman, A. Z., Ajit, A., & Weeds, Z. (2018). Classical and novel approaches to the analysis of honey and detection of adulterants. *Food Control*, 90, 152-165. http://dx.doi.org/10.1016/j.foodcont.2018.02.027.

Phillips, T., & Abdulla, W. (2022). A new honey adulteration detection approach using hyperspectral imaging and machine learning. *European Food Research and Technology*, 1-12. http://dx.doi.org/10.1007/s00217-022-04113-9. Online.

Pingzhen, W., Wenyong, W., & Shihai, Y. (2022). Research on consumers' perception of food risk based on LSTM sentiment classification. *Food Science and Technology*, 42, e47221. http://dx.doi.org/10.1590/fst.47221.

Pontes, M. J. C., Galvão, R. K. H., Araújo, M. C. U., Moreira, P. N. T., Pessoa, O. D. No., José, G. E., & Saldanha, T. C. B. (2005). The successive projections algorithm for spectral variable selection in classification problems. *Chemometrics and Intelligent Laboratory Systems*, 78(1-2), 11-18. http://dx.doi.org/10.1016/j.chemolab.2004.12.001.

Rasad, H., Entezari, M. H., Ghadiri, E., Mahaki, B., & Pahlavani, N. (2018). The effect of honey consumption compared with sucrose on lipid profile in young healthy subjects (randomized clinical trial). *Clinical Nutrition ESPEN*, 26, 8-12. http://dx.doi.org/10.1016/j.clnesp.2018.04.016. PMid:29908688.

Richards, L. E. (1988). Book review: Principal Component Analysis. *Journal of Marketing Research*, 25(4), 410.

Shamsudin, S., Selamat, J., Sanny, M., Bahari, S. A. R., Jambari, N. N., & Khatib, A. (2019). A comparative characterization of physicochemical and antioxidants properties of processed Heterotrigona itama honey from different origins and classification by chemometrics analysis. *Molecules*, 24(21), 3898. http://dx.doi.org/10.3390/molecules24213898. PMid:31671885.

Shao, Y. Y., Shi, Y. K., Xuan, G. T., Li, Q. K., Wang, F. H., Shi, C. K., & Hu, Z. C. (2022). Hyperspectral imaging for non-destructive detection of honey adulteration. *Vibrational Spectroscopy*, 118, 103340. http://dx.doi.org/10.1016/j.vibspec.2022.103340.

Shi, T., Chen, Y., Liu, H., Wang, J., & Wu, G. (2014). Soil organic carbon content estimation with laboratory-based visible-near-infrared reflectance spectroscopy: feature selection. *Applied Spectroscopy*, 68(8), 831-837. http://dx.doi.org/10.1366/13-07294. PMid:25061784.

Soares, S. F. C., Gomes, A. A., Araujo, M. C. U., Galvão, A. R. Fo., & Galvão, R. K. H. (2013). The successive projections algorithm. *Trends in Analytical Chemistry*, 42, 84-98. http://dx.doi.org/10.1016/j.trac.2012.09.006.

Song, S., Xiong, X., Wu, X., & Xue, Z. (2021). Modeling the SOFC by BP neural network algorithm. *International Journal of Hydrogen Energy*, 46(38), 20065-20077. http://dx.doi.org/10.1016/j.ijhydene.2021.03.132.

Sudershan, C. P., & Rao, S. V. N. N. (2020). Classification of crackle sounds using support vector machine. *Materials Today: Proceedings*. http://dx.doi.org/10.1016/j.matpr.2020.10.463. In press.

Wan, I., Hussin, N. N., Mazlan, S., Hussin, N. H., & Radzi, M. (2018). Physicochemical analysis, antioxidant and anti proliferation activities of honey, propolis and beebread harvested from stingless bee. *IOP Conference Series. Materials Science and Engineering*, 440(1), 012048.

Wang, L., Huang, Z., & Wang, R. (2021). Discrimination of cracked soybean seeds by near-infrared spectroscopy and random forest variable selection. *Infrared Physics & Technology*, 115, 103731. http://dx.doi.org/10.1016/j.infrared.2021.103731.

Wu, F. C., & Chyu, C. C. (2004). Optimization of correlated multiple quality characteristics robust design using principal component analysis. *Journal of Manufacturing Systems*, 23(2), 134-143. http://dx.doi.org/10.1016/S0278-6125(05)00005-1.

Wu, L., Du, B., Heyden, Y. V., Chen, L., Zhao, L., Wang, M., & Xue, X. (2016). Recent advancements in detecting sugar-based adulterants in honey – a challenge. *Trends in Analytical Chemistry*, 86, 25-38.

Yang, J., Wang, J., Lu, G., Fei, S., Yan, T., Zhang, C., Lu, X., Yu, Z., Li, W., & Tang, X. (2021). TeaNet: deep learning on Near-Infrared Spectroscopy (NIR) data for the assurance of tea quality. *Computers and Electronics in Agriculture*, 190, 106431. http://dx.doi.org/10.1016/j.compag.2021.106431.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X., & Zhuang, J. (2018). Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology*, 565, 720-736. http://dx.doi.org/10.1016/j.jhydrol.2018.08.050.

Zhang, G. Y., & Abdulla, W. (2022). New Zealand honey botanical origin classification with hyperspectral imaging. *Journal of Food Composition and Analysis*, 109, 104511. http://dx.doi.org/10.1016/j.jfca.2022.104511.

Zou, Z., Chen, J., Wang, L., Wu, W., Yu, T., Wang, Y., Zhao, Y., Huang, P., Liu, B., Zhou, M., Lin, P., & Xu, L. (2022a). Nondestructive detection of peanuts mildew based on hyperspectral image technology and machine learning algorithm. *Food Science and Technology*, 42, e71322. http://dx.doi.org/10.1590/fst.71322.

Zou, Z., Chen, J., Zhou, M., Wang, Z., Liu, K., Zhao, Y., Wang, Y., Wu, W., & Xu, L. (2022b). Identification of peanut storage period based on hyperspectral imaging technology. *Food Science and Technology*, 42, e65822. http://dx.doi.org/10.1590/fst.65822.

Zou, Z., Chen, J., Zhou, M., Zhao, Y., Long, T., Wu, Q., & Xu, L. (2022c). Prediction of peanut seed vigor based on hyperspectral images. *Food Science and Technology*, 42, e32822. http://dx.doi.org/10.1590/fst.32822.

Zou, Z., Long, T., Chen, J., Wang, L., Wu, X., Zou, B., & Xu, L. (2021). Rapid identification of adulterated safflower seed oil by use of hyperspectral spectroscopy. *Spectroscopy Letters*, 54(9), 675-684. http://dx.doi.org/10.1080/00387010.2021.1986543.

Zou, Z., Long, T., Wang, Q., Wang, L., Chen, J., Zou, B., & Xu, L. (2022d). Implementation of Apple's automatic sorting system based on machine learning. *Food Science and Technology*, 42, e24922. http://dx.doi.org/10.1590/fst.24922.

Zou, Z., Wu, Q., Wang, J., Xu, L., Zhou, M., Lu, Z., He, Y., Wang, Y., Liu, B., & Zhao, Y. (2023). Research on non-destructive testing of hotpot oil quality by fluorescence hyperspectral technology combined with machine learning. *Spectrochimica Acta. Part A: Molecular and Biomolecular Spectroscopy*, 284, 121785. http://dx.doi.org/10.1016/j.saa.2022.121785. PMid:36058172.