

Peaberry coffee discrimination using UV-visible spectroscopy combined with SIMCA and PLS-DA

By Diding Suhandy and Meinilwita Yulia



Peaberry coffee discrimination using UV-visible spectroscopy combined with SIMCA and PLS-DA

Diding Suhandy^a and Meinilwita Yulia^b

^aDepartment of Agricultural Engineering, Faculty of Agriculture, The University of Lampung, Bandar Lampung, Indonesia; ^bDepartment of Agricultural Technology, Lampung State Polytechnic, Rajabasa Bandar Lampung, Indonesia

ABSTRACT

Ultraviolet-visible (UV-Vis) spectral information (190–400 nm) was used to classify coffee samples into either pure peaberry or pure normal coffee classes using two chemometric methods: soft independent modelling of class analogy (SIMCA) and partial least squares discriminant analysis (PLS-DA). The spectral data of peaberry and normal coffee were acquired using a UV-Vis spectrometer (Genesys™ 10S UV-Vis, Thermo Scientific, USA). For both supervised discrimination methods, SIMCA and PLS-DA, all samples were correctly classified into their corresponding classes. The SIMCA model classified all samples accurately (100%) into either the peaberry or normal coffee class, even at a 5% confidence level; however, the PLS-DA model also correctly classified all samples (100%). Investigation of the major wavelengths contributing to the classifications using x-loading weights and loading of latent variables (LVs) indicated wavelengths at 230, 250, 270, 310, and 350 nm were important for determination of the coffee types. These wavelengths were closely related to the absorbance wavelengths of several important chemical components in roasted coffee: caffeine, caffeic acid, and chlorogenic acids (CGA). These results provided the basis for developing a simple and reliable method for peaberry coffee authentication (including more complex samples where peaberry coffee is blended with normal coffees) based on UV-Vis spectra.

ARTICLE HISTORY

Received 20 September 2016
Accepted 15 February 2017



KEYWORDS

Classification rate; Peaberry coffee; PLS-DA; SIMCA; UV-visible spectroscopy

Introduction

Coffee is one of the most popular beverages in the world,^[1] particularly in developed countries: with 4 and 5 kg consumed per capita in the US and Europe, respectively. On the other hand, the coffee beans are primarily grown in developing countries, with four such countries accounting for more than half of the world's production: Brazil, Vietnam, Colombia, and Indonesia.^[2]

In Indonesia, Sumatra Island dominates coffee cultivation, capturing 74.2% of the national production, with the largest production areas being Bengkulu, Lampung, and South Sumatra. Amongst the coffee types produced, there are several specialty coffees, which have a particularly high economic value, including civet coffee ('kopi luwak' in Indonesian language) and peaberry coffee ('kopi lanang' in Indonesian language). A peaberry (also called *caracol* or *snail* in Spanish) is a natural mutation of the coffee bean inside the cherry. Normally two coffee beans grow in a fruit (*dicotyledonous*)—flat against each other like halves of a peanut; however, on rare occasions a single bean is produced (*monocotyledon*). The production of peaberry coffee is very limited, with only

CONTACT Diding Suhandy  diding.sughandy@fp.unila.ac.id  Department of Agricultural Engineering, Faculty of Agriculture, The University of Lampung, Jl. Prof. Dr. Soemantri Brojonegoro No. 1 Bandar Lampung, 35145, Indonesia. Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ljfp

© 2017 Taylor & Francis Group, LLC

about 7% of any given coffee crop containing peaberry beans. The higher price for peaberry beans arises from its supposed more concentrated flavour compared with normal beans.

Recently, food authentication is becoming increasingly important due to the need to guarantee origin of the product and to assure no adulteration with contaminants or cheaper fillers has occurred.^[3] As a profitable trading product with the potential for adulteration and fraudulent trading, peaberry coffee is in need of such authentication procedures. At the coffee bean stage, even after roasting, peaberry (round in shape) and normal coffee beans (flat in shape) are easy to visually distinguish. However after grinding, discrimination of ground coffee made from peaberry or normal coffee cannot be visually done, nor can adulteration of the ground coffee be easily detected. Thus, in order to provide authentication of peaberry coffee and protect consumers from adulteration, a robust and easy method for identifying ground peaberry coffee and detection of its adulteration is needed.

Several methods have been used for the detection and quantification of coffee adulteration, as detailed in a recent study.^[4] Moreover, the feasibility of using spectroscopic methods (near-infrared and mid-infrared spectroscopy) for coffee classification and authentication has been investigated in previous studies as well.^[5–7] Recently, ultraviolet-visible (UV-Vis) spectroscopy has been applied to the detection and quantification of normal coffee adulteration with coffee husks.^[8,9] This method has also been used for discrimination between normal civet coffee and normal non-civet coffee.^[10,11] The advantages of this technique include it is inexpensive and is a routinely used analytical technique, characteristics that lend itself to application in Indonesia. However, to date there has been no report on the use of this method for discrimination between pure ground peaberry and normal coffee. Therefore in this research, we investigate the use of UV-Vis spectra combined with chemometrics methods (soft independent modelling of class analogy, SIMCA; and partial least squares discriminant analysis, PLS-DA) to discriminate between ground peaberry and normal coffee samples.

Material and method

Sample preparation

A 1 kg sample of roasted peaberry coffee beans (robusta variety: *Coffea canephora*) was collected directly from a coffee farmer at Liwa, Lampung, Indonesia (Hasti coffee Lampung), while another 1 kg sample of roasted normal coffee beans (robusta variety) was also collected for making a normal (non-peaberry) ground coffee sample. All coffee beans were harvested manually from the same plantation and manually separated into peaberry and normal coffee beans. These coffee samples were then separately ground using a home-coffee grinder (Sayota). Since particle size in coffee powder has a significant influence on the spectra obtained,^[12] we sieved both coffee sample types through a nest of U.S. standard sieves (mesh number 40) on a Meinzer II sieve shaker (CSC Scientific Company, Inc. USA) for 10 min to obtain a particle size of 420 μm . The experimental measurements were performed at room temperature (around 27–29°C). In this research we prepared 50 samples each of ground peaberry and normal coffee.

For developing and evaluating the calibration model, the samples were divided into two groups: a calibration and a prediction sample set. The calibration sample set had 68 samples (34 samples each of peaberry and normal coffee) that were used for developing the calibration model with full cross-validation. The prediction sample set had 32 samples (16 samples each of peaberry and normal coffee) that were used for evaluating the performance of the developed calibration model.

A standard aqueous extract of the coffee samples was then obtained.^[8,9] First, 1.0 g of each sample was weighed and placed in a glass beaker. Then 10 mL of distilled water at 90–98°C was added and mixed with a magnetic stirrer (Cimarec™ Stirrers, model S130810-33, Barnstead International, USA) at 350 rpm for 5 min. Then the samples were filtered using a 25-mm pore-sized quantitative filter paper coupled with an Erlenmeyer flask. After cooling to room temperature (for 20 min), all extracts were then diluted in a 1:20 (mL: mL) proportion with distilled water.

Spectral data acquisition using a UV-Vis spectrometer

Spectral data of these diluted peaberry and normal coffee extracts were acquired using a UV-Vis spectrometer (Genesys™ 10S UV-Vis, Thermo Scientific, USA) in the range of 190–700 nm. This spectrometer is a dual-beam spectrometer equipped with a five-cell sample holder and one cell for a blank or reference. The wavelength accuracy is 1 nm, with dual silicon photodiodes as the detector and a Xenon flash as the illumination source.

The absorbance data of peaberry and normal coffee extracts were collected at room temperature. For this, we put 3 mL of extract into the cuvettes with a 10 mm path length. Before measuring the samples, a blank (the same distilled water used in the extraction process) was placed inside the reference cell holder. In this research, relative absorbance spectra were calculated as follows:

$$A_{\lambda} = -\log_{10} \frac{S_{\lambda}}{R_{\lambda}}$$

where S_{λ} is the intensity of the sample at wavelength λ and R_{λ} is the intensity of the reference at the same wavelength. The “relative absorbance spectra”, hereafter referred to as “spectra”, were transferred to a computer for further multivariate analysis. Spectral preprocessing (principal component analysis (PCA), PLS-DA, or SIMCA) was carried out using Unscrambler 9.7 (CAMO Software AS, Norway).

Analysis data using the SIMCA method

SIMCA is a method of supervised data classification based on a PCA methodology. The theory of SIMCA has been extensively discussed and can be found elsewhere.^[13–15] First, a PCA model was created for each class using the calibration data set for each corresponding class. In SIMCA modelling, the number of PCAs is equal to the number of classes of samples, because each class needs to create its own PCA model for making the SIMCA model for each class. For this, we developed a SIMCA model for both the peaberry and normal coffee samples. A critical point in developing the SIMCA model is to determine the optimal number of principal components (PCs) involved in the model. Here, we used a full cross-validation method and calculated the root mean square error of cross-validation (RMSECV). The lowest RMSECV value corresponds to the optimal number of PCs. The performance of the SIMCA model was then evaluated using the prediction sample set. Samples in the prediction set are projected into each SIMCA model and their residual distance calculated. Residual standard deviation was used as a measure of the critical distance for classification. If the residual distance from the model was below their statistical limit, the sample was defined to that class and if the residual variance was higher, then the sample did not belong to that class.

Analysis data using PLS-DA

PLS-DA is based on a PLS regression algorithm (in this study we used a PLS1 algorithm with only one dependent Y variable),^[16] which searches for latent variables (LVs) with a maximum covariance with the Y-variables. PLS-DA attempts to build models that can maximize the separation among the classes of objects. Detailed explanations about the PLS-DA algorithm can be found in the literature.^[17–20] In our PLS-DA, each sample in the calibration set was assigned a dummy variable as a reference value (variable y), which is an arbitrary number designating whether the sample belongs to a particular class or not^[21] (1 = peaberry coffee; 0 = normal coffee). The PLS-DA calibration model was developed using the calibration sample set and the optimal number of LVs was determined using the RMSECV value. The lowest RMSECV value corresponds to the optimal number of LVs. The performance of the PLS-DA model was evaluated using the prediction sample set using ± 0.5 as a threshold value to delimit the classes. A sample was considered to be correctly classified if the predicted value lay on the same side of the midpoint of the assigned value of 0.5, a cut-off criterion that is often used.^[22–25] A coffee sample was classified as peaberry coffee if its value was above 0.5 and classified as normal coffee if the value was below 0.5.

Results and discussion

Spectral data analysis of peaberry and normal coffee samples

Figure 1 shows the spectral data of peaberry and normal coffee samples in the range of 190–700 nm (ultraviolet and visible region). It can be seen that the spectra of both coffees have a similar shape but not the same absorbance intensity. The spectra of normal coffee have higher absorbance intensity, especially in the ultraviolet region (after 390 nm). In the visible region, both spectra have almost same absorbance intensity and overlapped each other. However, beyond 290 nm the peaberry and normal coffee spectra often overlap and it is not easy to discriminate between the two. This spectral data (190–400 nm) will be used for further analysis.

In the range 190–400 nm, several peaks and troughs can be observed, including peaks at 281 and 320 nm. The troughs are at 256 and 310 nm. These peaks and troughs are closely related to the absorbance of several important chemical components in roasting coffee. For example, the peak at 281 nm is closely related to the absorbance of caffeine, while those at 310 nm and 320 nm are related to the absorbance of caffeic acid.^[8] The trough at 256 nm is closely related to the absorbance of vanillic acid.

Principal components analysis

First, a PCA was performed on the whole dataset (100 samples) using original, as well as preprocessed spectral data in the range 190–400 nm. Preprocessing of the spectra using the Savitzky–Golay 1st derivative algorithm resulted in a very clear separation between peaberry and normal coffee samples. Figure 2 shows the PCA scores along the first two PCs, which represent the quasi-totality of the total variance (94%). Figure 2 shows that a clear class separation was attained, where two distinct clusters were observed separated by the PC1 axis. Peaberry coffee samples are clearly clustered on the left while normal coffee samples appeared on the right of the PC1 axis.

Classification result using the SIMCA method

SIMCA classification for discrimination between peaberry and normal coffee samples was developed by generating the individual PCA models for each class of coffee samples (peaberry and normal) using the calibration sample set (34 samples for each class). The SIMCA model of peaberry and normal coffee was then used to predict the type of coffee samples in the prediction sample set (16 samples from each class). The prediction results showed that the SIMCA model accurately classified

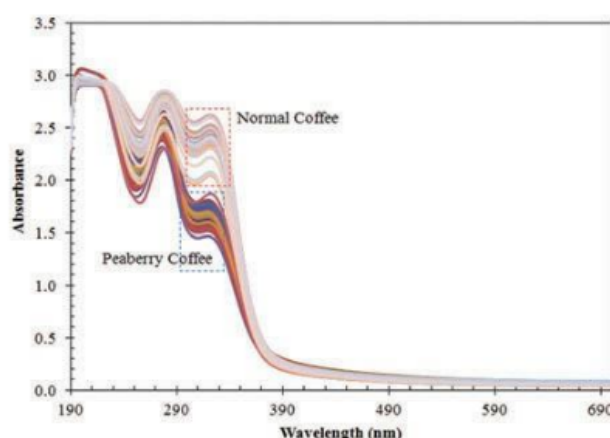


Figure 1. Original spectra of peaberry and normal coffee samples in ultraviolet-visible region (190–700 nm).

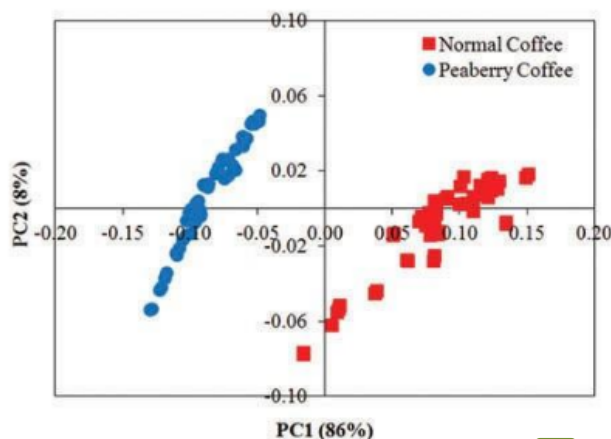


Figure 2. Principal components analysis of spectra from the peaberry and normal coffee samples (projection onto the first two principal components).

(100%) all of the samples, either as the peaberry or normal coffee, down to a significance level of 5% level, as shown in Fig. 3. From the figure, it can be seen that all samples were separated and closely clustered to the class that they should belong to. As can be seen, no samples were located in the lower left quadrant, which would indicate that they belonged to both classes simultaneously. There were also no samples in the upper right quadrant, which indicates a sample that does not belong to either of the defined models. The upper left and lower right quadrants define samples that belong to a particular class.

To calculate the performance of the classification, a confusion matrix was created, with the accuracy, sensitivity, and specificity of the coffee sample classifications using the developed SIMCA models (Table 1). It can be seen that prediction results (using 16 samples for each class) were satisfactory (100%) for the accuracy, sensitivity, and specificity rates.

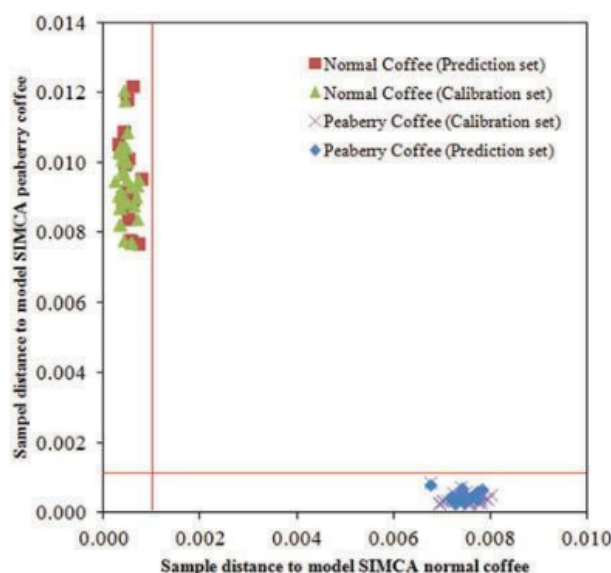


Figure 3. Cooman plot for the classification of peaberry and normal coffee samples (vertical and horizontal red lines show 95% confidence interval).

6

Table 1. Confusion matrix with accuracy, sensitivity, and specificity of the classification of coffee samples by SIMCA using the prediction sample set.

	Class peaberry coffee (assigned by the SIMCA model peaberry)	Class normal coffee (assigned by the SIMCA model normal)
Class peaberry coffee (actual)	16	0
Class normal coffee (actual)	0	16
Accuracy (%)	100	
Sensitivity (%)	100	
Specificity (%)	100	

Classification result using the PLS-DA method

3

A supervised discriminant method, PLS-DA, was carried out to develop a calibration model between spectral data (as predictor variables) and the type of coffee (as target variables). This PLS-DA model was used to predict the type of coffee in the prediction sample set and the results are depicted in Fig. 4. It can be clearly seen that both peaberry and normal coffee samples were limited to their own class. For example, all peaberry samples (with the o mark) had a value above 0.5 and below 1.5 and therefore all peaberry samples correctly classified to be 1: peaberry coffee. All normal coffee samples (with the x mark) had a value above -0.5 and below 0.5 and therefore all normal coffee samples were correctly classified to be 0: normal coffee. From this we can see that the prediction results using the PLS-DA model were satisfactory with a 100% correct classification rate for peaberry and normal coffee samples.

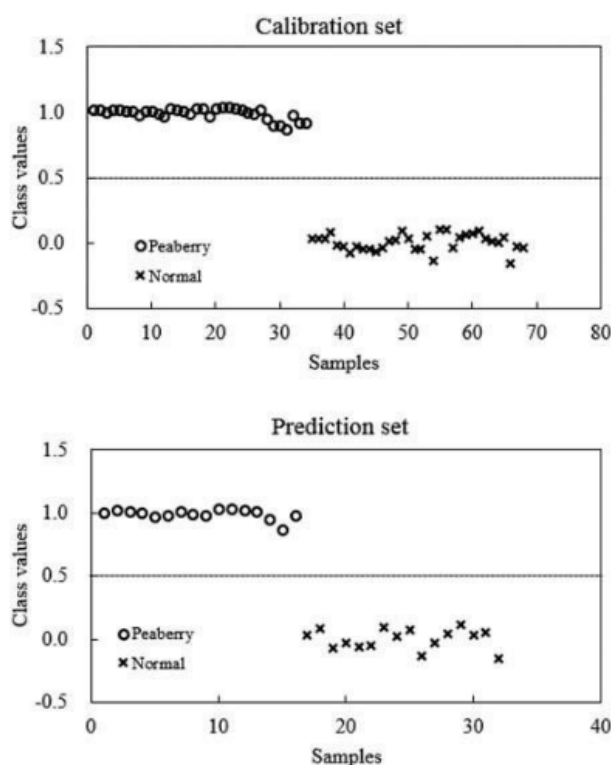


Figure 4. Estimated class values by the PLS-DA model versus the calibration and prediction sample plots for discrimination between peaberry and normal coffees.

Investigation on sensitive wavelengths

Both the SIMCA and PLS-DA models involved the use of a number of predictor variables (210 wavelength variables in the range from 190 to 400 nm). It was important to identify the major wavelengths that drive these models. For this we plotted the x-loading results (from SIMCA) and loading of LVs (from PLS-DA). The x-loadings indicate how well the x-variable (wavelength) is taken into account by the model components. It can be used to understand how much each of the x-variables (wavelengths) contributed to meaningful variation in the data, and to interpret variable relationships. It is also useful for interpreting the meaning of each model component. The loading weights indicate how much each wavelength (x-variable) contributes to explaining the response variation (type of coffee) along each model component. The loading weights were normalized, so that their length could be interpreted, as well as their direction. Wavelengths (x-variables) with large loading weight values are important for the determination of the type of coffee (peaberry or normal coffee). With a similar function, loading of LVs is primarily used to check the influence of wavelength (x-variable) on determining the type of coffee (y-variable). A large absolute value indicates a significant influence of wavelength.

From Fig. 5 several wavelengths can be identified with a high x-loading weight, including wavelengths at 230, 270, and 350 nm. In Fig. 6 several wavelengths with high loading of LVs at 230, 250, 270, 310, and 350 nm can be observed. Those wavelengths are considered to be important for the differentiation between the classes. These wavelengths are closely related to the absorbance of several important chemical components in roasted coffee. For example, the wavelength at 270 nm is related to the absorbance of caffeine, while those at 310 and 350 nm are closely related to the absorbance of caffeic acid and CGA, respectively.^[8,26]

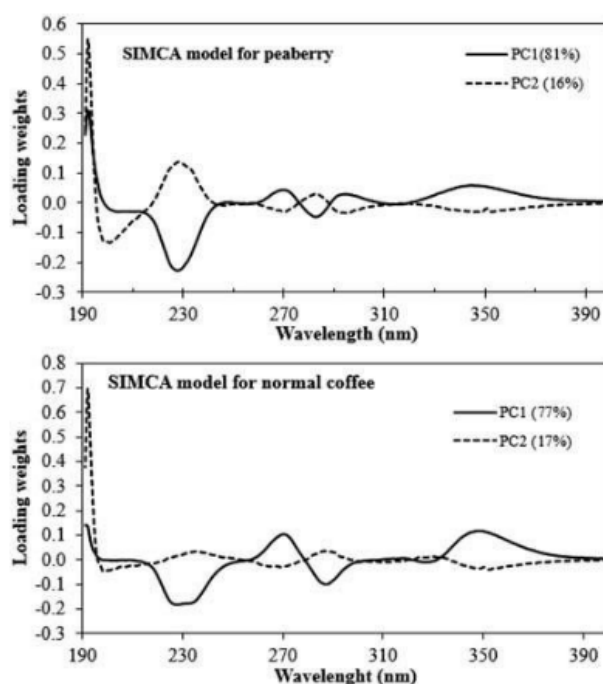


Figure 5. X-Loading weights versus wavelength of the top two PCs (PC1 and PC2) of coffee samples (upper: peaberry coffee; lower: normal coffee).

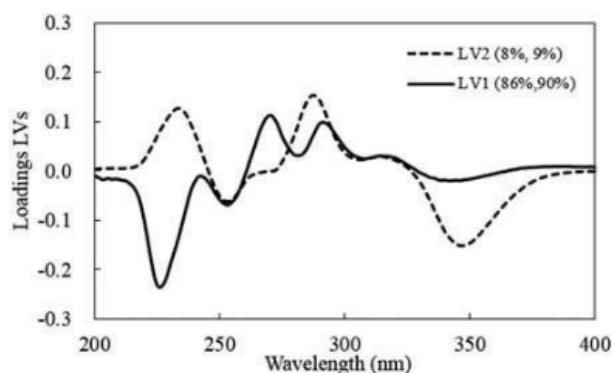


Figure 6. Plots of loadings of latent variables versus wavelength for the PLS-DA model.

Conclusion

This research demonstrates the potential of UV-Vis spectral information to be used for robust and simple discrimination of ground pure peaberry and ground normal coffee samples. The two supervised discrimination methods investigated, SIMCA and PLS-DA, provided satisfactory classification rates, with all samples being correctly classified into their corresponding classes. Subsequent investigation of the particularly sensitive wavelengths using x-loading weights and loading of LVs showed that absorbance intensities at 230, 250, 270, 310, and 350 nm are important for discriminating between the types of coffee examined in this study. Those wavelengths are closely related to the absorbance wavelength of several important chemical components in the roasted coffee, such as caffeine, caffeic acid, and CGA.

Acknowledgements

We would like to thank Department of Agricultural Engineering, The University of Lampung for allowing the use of the UV-visible spectrometer and its supporting devices. We also would like to thank Hasti Coffee Lampung for providing the coffee samples. We also acknowledge Professor Garry John Piller (Graduate School of Agriculture, Kyoto University, Japan) for his help and useful discussions during the preparation of the manuscript.

ORCID

Diding Suhandy  <http://orcid.org/0000-0001-5555-1887>

Meinilwita Yulia  <http://orcid.org/0000-0002-5197-2796>

Peaberry coffee discrimination using UV-visible spectroscopy combined with SIMCA and PLS-DA

ORIGINALITY REPORT

20%

SIMILARITY INDEX

PRIMARY SOURCES

- | | | |
|----------|---|-----------------------|
| 1 | worldwidescience.org
<small>Internet</small> | 162 words — 5% |
| <hr/> | | |
| 2 | Liu, F.. "Classification of brands of instant noodles using Vis/NIR spectroscopy and chemometrics", Food Research International, 2008
<small>Crossref</small> | 96 words — 3% |
| <hr/> | | |
| 3 | Abdul-Rahman A. Roshan, Haidy A. Gad, Sherweit H. El-Ahmady, Mohamed I. Abou-Shoer, Mohamed S. Khanbash, Mohamed M. Al-Azizi. "Characterization and Discrimination of the Floral Origin of Sidr Honey by Physicochemical Data Combined with Multivariate Analysis", Food Analytical Methods, 2016
<small>Crossref</small> | 66 words — 2% |
| <hr/> | | |
| 4 | Cozzolino, D.. "Can spectroscopy geographically classify Sauvignon Blanc wines from Australia and New Zealand?", Food Chemistry, 20110515
<small>Crossref</small> | 60 words — 2% |
| <hr/> | | |
| 5 | Anisur Rahman, Naoshi Kondo, Yuichi Ogawa, Tetsuhito Suzuki, Yuri Shirataki, Yumi Wakita. "Classification of fresh and spoiled Japanese dace (Tribolodon hakonensis) fish using ultraviolet-visible spectra of eye fluid with multivariate analysis", Engineering in Agriculture, Environment and Food, 2016
<small>Crossref</small> | 40 words — 1% |

-
- 6 Urijatan Teixeira de Carvalho Polar Souto, Mayara Ferreira Barbosa, Hebertty Vieira Dantas, Aline Santos de Pontes et al. "Screening for Coffee Adulteration Using Digital Images and SPA-LDA", Food Analytical Methods, 2014
Crossref 39 words — 1%
-
- 7 www.condaxis.com
Internet 32 words — 1%
-
- 8 www.science.gov
Internet 30 words — 1%
-
- 9 Yi Chen, Jing Deng, Yuanxing Wang, Boping Liu, Jian Ding, Xuejin Mao, Juan Zhang, Haitao Hu, Jing Li. "Study on discrimination of white tea and albino tea based on near-infrared spectroscopy and chemometrics", Journal of the Science of Food and Agriculture, 2014
Crossref 28 words — 1%
-
- 10 Liu, Xiaona, Qiao Zhang, Zhisheng Wu, Xinyuan Shi, Na Zhao, and Yanjiang Qiao. "Rapid Elemental Analysis and Provenance Study of Blumea balsamifera DC Using Laser-Induced Breakdown Spectroscopy", Sensors, 2014.
Crossref 25 words — 1%
-
- 11 Barbosa, Mayara F., Hebertty V. Dantas, Pablo N. T. Moreira, Roberto K. H. Galvão, and Mário C. U. Araújo. "Use of an Automatic System in the Preparation of Gas Mixtures for Multivariate Calibration: A Case Study Involving NIR Analysis of Natural Gas", Journal of the Brazilian Chemical Society, 2015.
Crossref 16 words — < 1%
-
- 12 Castro-Suarez, John R., Migdalia Hidalgo-Santiago, and Samuel P. Hernández-Rivera. "Detection of Highly Energetic Materials on Non-Reflective

- 13 Fei Liu, Yong He, Li Wang. "Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis", *Analytica Chimica Acta*, 2008 15 words — < 1%

- 14 P PERALTAZAMORA, L CORNEJOPONCE, N NAGATA, R POPPI. "Chemometric alternatives for resolution of classical analytical problemsSpectrophotometric determination of lanthanide mixtures", *Talanta*, 1997 15 words — < 1%

- 15 Santos, Maria B. H., Adriano A. Gomes, Welma T. S. Vilar, Pollyne B. A. Almeida, Máira Milani, Márcia B. M. Nóbrega, Everaldo P. Medeiros, Roberto K. H. Galvão, and Mário C. U. Araújo. "Non-Destructive NIR Spectrometric Cultivar Discrimination of Castor Seeds Resulting from Breeding Programs", *Journal of the Brazilian Chemical Society*, 2014. 10 words — < 1%

- 16 Diding Suhandy, Meinilwita Yulia, Yuichi Ogawa, Naoshi Kondo. "Prediction of vitamin C using FTIR-ATR terahertz spectroscopy combined with interval partial least squares (iPLS) regression", 2011 IEEE/SICE International Symposium on System Integration (SII), 2011 10 words — < 1%

- 17 Ioannis S. Arvanitoyannis, Antonios Vlachos. "Implementation of Physicochemical and Sensory Analysis in Conjunction with Multivariate analysis towards Assessing Olive Oil Authentication/Adulteration", *Critical Reviews in Food Science and Nutrition*, 2007 8 words — < 1%

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON

EXCLUDE MATCHES OFF