

The classification of arabica gayo wine coffee using UV-visible spectroscopy and PCA-DA method

By Diding Suhandy and Meinilwita Yulia

The classification of arabica gayo wine coffee using UV-visible spectroscopy and PCA-DA method

Diding Suhandy^{1,*}, and Meirilwita Yulia²

¹The University of Lampung, Faculty of Agriculture, Department of Agricultural Engineering, Spectroscopy Research Group (SRG), Laboratory of Bioprocess and Postharvest Engineering, Jl. Prof. Dr. Soemantri Brojonegoro No.1, Bandar Lampung, 35145, Indonesia

²Lampung State Polytechnic, Department of Agricultural Technology, Jl. Soekarno Hatta No. 10, Rajabasa Bandar Lampung, 35141, Indonesia

Abstract. The unique processing of Arabica Gayo Wine coffee produces special attributes to the beverage and could increase its value. However, it is important to prove the authenticity of Arabica Gayo Wine coffee using reliable methods. The objective of this study was to evaluate the potential of UV-visible spectroscopy and principal component analysis-discriminant analysis (PCA-DA) method for classification of ground roasted Arabica Gayo Wine coffee. A number of 200 samples of Arabica Gayo Wine coffee and 200 samples of Arabica Gayo normal (not Wine) coffee was used. The spectral data obtained in the UV-visible region were analyzed using PCA-DA with standard normal variate (SNV) and followed by Savitzky-Golay smoothing with different number of smoothing point (NSP). The results showed that the best PCA-DA model was obtained with NSP = 23 with coefficient of determination for calibration (R^2) = 0.99, root mean square error of calibration (RMSEC) = 0.005692 and root mean square error of validation (RMSEV) = 0.006112. Using this model, a good classification between Gayo Wine and Gayo normal in prediction step was achieved with 100% accuracy, sensitivity and specificity. Thus, the proposed method can be used for the evaluation of authenticity of ground roasted Arabica Gayo Wine coffee.

1 Introduction

Gayo Arabica coffee from Aceh, Indonesia is one of the important and most expensive specialty coffees in Indonesia. In 2016, the total production of Gayo Arabica coffee in Aceh province (central Aceh, Bener Meriah and Gayo Lues region) is about 47 thousand tons (total national coffee production is about 639 thousand tons) [1]. Now, Aceh is the second largest production area for Arabica coffee in Indonesia [1-2].

Recently, a special production of Gayo Arabica coffee from Takengon, Aceh with several time of fermentation called as Gayo Wine coffee is becoming popular. This coffee has a unique taste and high in price. The popularity and limited supply of Gayo Wine coffee makes it very vulnerable to fraudulent [3].

In practice, it is not easy to discriminate ground roasted Gayo Wine and Gayo normal (not Wine) coffee by naked eyes. As shown in Fig. 1, visible inspection of the two coffee is not reliable. For this reason, it is desired to develop a new analytical method to discriminate Gayo Wine and Gayo normal in groundroasted coffee. Among the available and possible analytical method, UV-visible spectroscopy is preferable due to its several advantages: low cost spectrometer, easy to use and to do maintenance of spectrometer, free- chemical waste and the spectrometer is available and accessible in most laboratory in Indonesia [4].

In previous studies, UV-visible spectroscopy has been used to discriminate ground roasted pure peaberry coffee and pure normal coffee [4], to discriminate ground roasted civet and non-civet coffee [5], and to quantify the degree of adulteration in ground roasted civet coffee [6]. Discrimination of several Indonesian specialty coffees has been investigated using Fluorescence spectroscopy [7] and NIR spectroscopy [8]. However, to the best of our knowledge, there is no any report on the use of UV-visible spectroscopy for discrimination ground roasted Gayo Wine and Gayo normal coffee. Therefore, the objective of this study is to evaluate the potential of UV-visible spectroscopy combined with chemometric method to discriminate pure ground roasted Gayo Wine and Gayo normal coffee.



Fig. 1. Ground roasted Arabica Gayo Wine and Gayo normal.

* Corresponding author: diding.sughandy@fp.unila.ac.id

2 Materials and method

2.1 Coffee samples

Ground roasted Arabica Gayo Wine and Gayo normal coffee were purchased from local market in Banda Aceh, Indonesia and transported to Bandar Lampung. A total of 200 pure Gayo Wine and 200 pure Gayo normal samples were used. Each sample has 1 gram pure Gayo Wine or Gayo normal ground roasted coffee. Sample preparation including sieving and grinding was done based on previous reported studies [4-6].

Spectral data acquisition was done on aqueous coffee samples. For this, an extraction of coffee samples were performed based on previous researches [4-6].

2.2 Spectral measurement

UV-visible spectral data measurements were performed with a Genesys 10s UV-Vis Spectrometer (Thermo Fisher Scientific, USA) equipped with a 10 mm quartz cell. Spectral measurements were recorded in the range of 190-1100 nm with 1 nm resolution. Distilled water was used for blank measurement.

2.3 Chemometrics

Prior to classification analysis (both unsupervised and supervised methods), spectra pre-processing was performed. There were two pre-processing methods applied to improve the quality of spectral data: standard normal variate (SNV) and Savitzky-Golay (SG) smoothing with different number of smoothing point (NSP). SNV is used to remove both an additive and multiplicative noise due to scatter interference [9-11]. SNV has been used for many spectrometric application and can improve the prediction accuracy [9]. In practice, SNV was done by subtracting the mean from an individual spectrum and divides it by its standard deviation [12]. SG smoothing is a widely used pre-processing method that can effectively eliminate noises like baseline-drift, tilt and reverse [10].

Principal component analysis (PCA) is one of the unsupervised methods used for dimensionality reduction by converting a number of correlated variables into fewer variables called principal components (PCs). In this study, PCA was used for clustering of samples and detecting the occurrence of outliers. Classification model to discriminate between Gayo Wine and Gayo normal was developed by using principal component analysis-discriminant analysis PCA-DA. This method worked based on principal component regression (PCR) algorithm with a dummy response variables ($y=1$ for Gayo Wine and $y=0$ for Gayo normal). For more detail about PCA and PCR, several previous published paper were suggested [13-14]. For PCA-DA model development, samples were divided randomly into three sets: calibration samples set (232 samples), validation sample set (114 samples) and prediction sample set (56 samples). To evaluate the classification performance of the PCA-DA model, three commonly parameters were

used: overall accuracy, sensitivity and specificity. The calculation of those parameters was done based on previous published work [11].

2.4 Software

The standard multivariate analysis package, The Unscrambler X version 10.4 (free trial version, 30 days, CAMO, Oslo, Norway) was used for the data analysis including pre-processing spectra, PCA and PCA-DA.

3 Results and discussion

3.1 Analysis of spectral data

Figure 2 shows the average spectra of Gayo Wine and Gayo normal coffee samples in the range of 190-1100 nm. The most informative spectral region was identified in the range of 250-450 nm. In the initiation of spectral data, a relatively noisy data was observed and spectral region more than 450 nm is almost no absorbance could be detected (almost zero absorbance). Figure 3 shows the average pre-processed spectra in the range of 250-450 nm. The pre-processing methods are combination of standard normal variate and Savitzky-Golay smoothing (NSP=3). It is also said that a spectral difference was observed in the region of 250-450 nm.

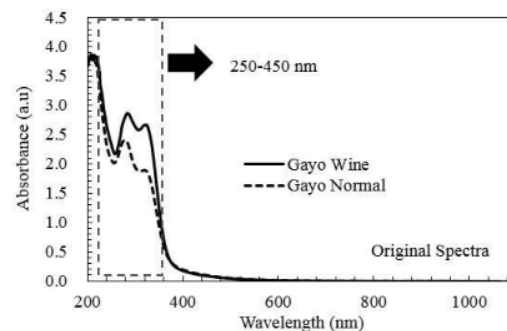


Fig. 2. Original spectra of Gayo Wine and Gayo normal (average) in the range of 190-1100 nm.

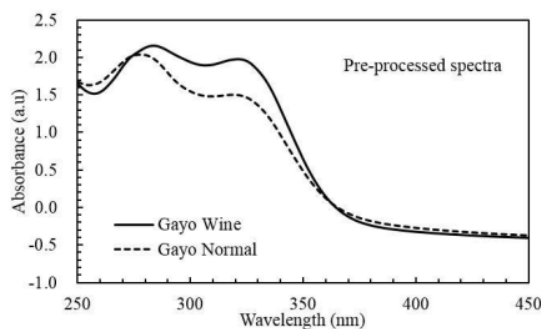


Fig. 3. Pre-processed spectra of Gayo Wine and Gayo normal (average) in the range of 250-450 nm.

3.2 Principal component analysis (PCA)

In order to see the clustering of all samples, PCA was performed for 400 samples (200 samples Gayo Wine and 200 samples Gayo normal) with full cross-validation method using spectral data 250-450 nm. The result of PCA was depicted in Fig. 4. The score plot of the first two principal components (PCs) clearly shows that Gayo Wine and Gayo normal samples were well-separated along PC1: Gayo normal samples have negative PC1 scores, whereas Gayo Wine samples have positive PC1 scores. The PC1 scores alone accounted for more than 90% of the spectral data and the cumulative percent variance (CPV) of the first two principal components (PC1+PC2) was found to be 97% of the total variance in the spectral datasets.

To determine the important wavelengths which is responsible for the separation between Gayo Wine and Gayo normal coffee samples, a plot of x-loadings versus wavelengths was investigated and the result was showed in Fig. 5. Wavelengths with high loading values are the most important variables for separation between Gayo Wine and Gayo normal coffee. Two wavelengths at 262 nm and 325 nm were observed with high negative and positive loadings, respectively. Those wavelengths were correspond with the absorbance of caffeine and trigonelline in coffee, respectively [4-6].

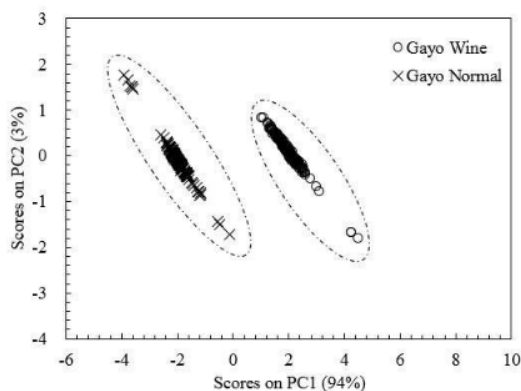


Fig. 4. Principal component analysis plot of PC1 and PC2 of Gayo Wine and Gayo normal in the range of 250-450 nm.

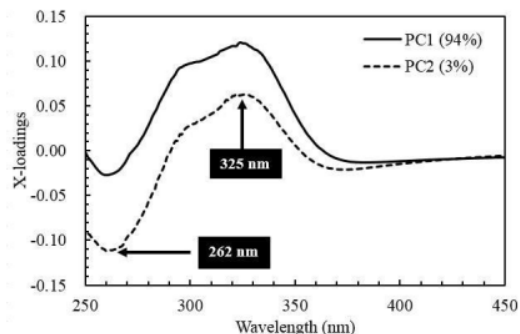


Fig. 5. The x-loadings versus wavelength of first two principal components (PC1 and PC2) in the range of 250-450 nm.

3.3 Development of PCA-DA classification model

The PCA-DA model was developed using pre-processed spectra (combination of SNV+SG smoothing). In the SG smoothing, there are many different parameters include polynomial degree (PD), the derivative order of polynomial (DOP) and the number of smoothing points (NSP) or the size of smoothing window (SW) [15]. A too-small NSP is prone to cause calculation error, resulting in a decreased model precision, while a too-large NSP would over smooth the spectral data, most of the information were missing and leading to decreased accuracy of model [15]. For this reason, it is important to select an appropriate NSP for SG smoothing. Figure 6 shows the plot between NSP and root mean square error of validation (RMSEV). The NSP = 23 was corresponding with the lowest RMSEV = 0.006112. This PCA-DA model has 15 latent variables (LVs), high coefficient of determination ($R^2 = 0.99$) and low root mean square error of calibration (RMSEC = 0.005692). This PCA-DA model with NSP = 23 was selected for the best classification model to discriminate Gayo Wine and Gayo normal ground roasted coffee.

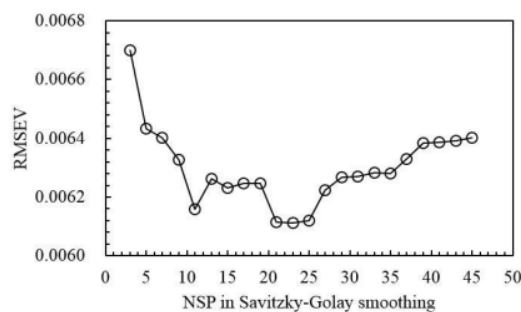


Fig. 6. Plot of NSP versus RMSECV in validation step.

The best PCA-DA model was presented in plot of samples versus y-predicted values both for calibration (Fig. 7) and validation (Fig. 8) ($y=1$ for Gayo Wine and $y=0$ for Gayo normal). The threshold value is 0.5 line. It shows that all samples in calibration and validation are located very close to its class resulted in 100% of accuracy, sensitivity and specificity.

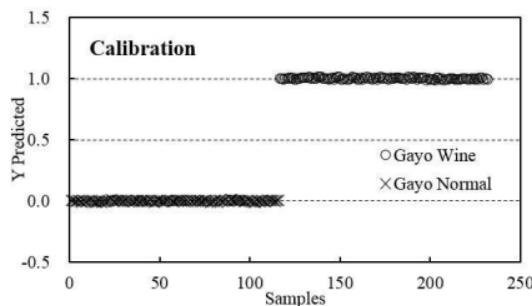


Fig. 7. PCA-DA classification model in calibration step.

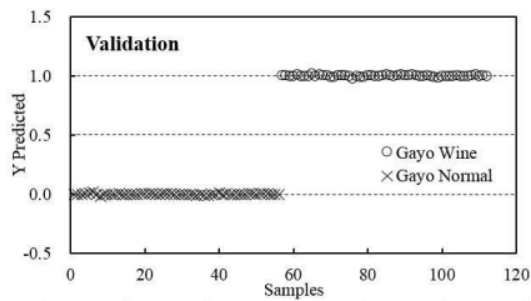


Fig. 8. PCA-DA classification model in validation step.

Our result was comparable to previous reported studies. Marquetti *et al.* studied the discrimination of several Arabica coffees based on its geographic and genotypic origin using NIR spectroscopy. The sensitivity was 75-100% for geographic and genotypic discrimination [16]. Suhandy and his co-workers has developed model discrimination for several Indonesian specialty using UV-visible spectroscopy and chemometric methods with 100% of accuracy, sensitivity, and specificity: peaberry vs. normal coffee [4] and civet vs. non-civet coffee [5].

3.4 Blind classification using PCA-DA model

To evaluate the predictive ability of the developed PCA-DA model in classification of Gayo Wine and Gayo normal, 56 samples (28 samples for Gayo wine and 28 samples for Gayo normal, respectively) which is not used in model development was tested. The result was depicted in Fig. 9. It is clear that all prediction samples are lied in its proper class. All Gayo normal samples were close to $y=0$ and all Gayo Wine samples were close to $y=1$. Thus, the accuracy, sensitivity and specificity of 100% was also obtained in prediction.

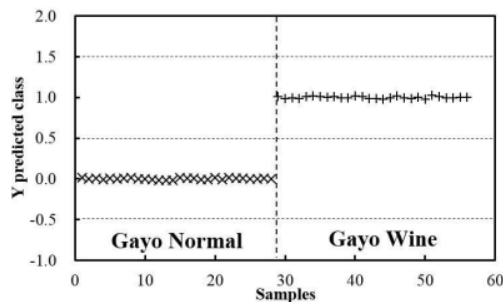


Fig. 9. PCA-DA classification result using prediction set.

4 Conclusions

This study has revealed that UV-visible spectroscopy combined with PCA-DA method has potential for discrimination of Gayo Wine and Gayo normal ground roasted coffee. The developed PCA-DA model is able to predict the class of Gayo Wine and Gayo normal with

100% of accuracy, sensitivity and specificity. This method can be implemented as simple, consistent and green technology to allow us to discriminate Gayo Wine and Gayo normal. In the near future, this kind of technology is needed to develop an authentication system of Indonesian specialty coffee.

The authors thank the University of Lampung and the Indonesian Ministry of Research, Technology and Higher Education (Kemristekdikti) for funding this research under Research Grant PENELITIAN STRATEGIS NASIONAL INSTITUSI (PSNI) [grant No. 393/UN26.21/PN/2018].

The classification of arabica gayo wine coffee using UV-visible spectroscopy and PCA-DA method

ORIGINALITY REPORT

12%

SIMILARITY INDEX

PRIMARY SOURCES

1 M Yulia, D Suhandy. "Indonesian palm civet coffee discrimination using UV-visible spectroscopy and several chemometrics methods", *Journal of Physics: Conference Series*, 2017 59 words — 3%

[Crossref](#)

2 Giuseppe Toscano, Åsmund Rinnan, Andrea Pizzi, Manuela Mancini. "The Use of Near-Infrared (NIR) Spectroscopy and Principal Component Analysis (PCA) To Discriminate Bark and Wood of the Most Common Species of the Pellet Sector", *Energy & Fuels*, 2017 39 words — 2%

[Crossref](#)

3 Pedro Sousa Sampaio, Andreia Soares, Ana Castanho, Ana Sofia Almeida, Jorge Oliveira, Carla Brites. "Optimization of rice amylose determination by NIR-spectroscopy using PLS chemometrics algorithms", *Food Chemistry*, 2018 34 words — 2%

[Crossref](#)

4 Meinilwita Yulia, Diding Suhandy. "The qualitative and quantitative analysis of adulteration in specialty coffee from Tanggamus lampung using UV-visible spectroscopy and chemometrics", AIP Publishing, 2021 34 words — 2%

[Crossref](#)

-
- 5 "EMBEC & NBC 2017", Springer Science and Business Media LLC, 2018 20 words — 1%
Crossref
-
- 6 www.scilit.net 14 words — 1%
Internet
-
- 7 op.niscair.res.in 13 words — 1%
Internet
-
- 8 www.atomtrace.com 12 words — 1%
Internet
-
- 9 Kalivas, J.H.. "Pareto calibration with built-in wavelength selection", *Analytica Chimica Acta*, 20040303 9 words — < 1%
Crossref
-
- 10 sciex.com 9 words — < 1%
Internet
-
- 11 Edwin R. Arboleda. "Discrimination of civet coffee using near infrared spectroscopy and artificial neural network", *International Journal of Advanced Computer Research*, 2018 9 words — < 1%
Crossref
-
- 12 Marcelo A. Jurado Zavaleta, Mirta R. Alcaraz, Lidia G. Peñaloza, Analía Boemo et al. "Chemometric modeling for spatiotemporal characterization and self-depuration monitoring of surface water assessing the pollution sources impact of northern Argentina rivers", *Microchemical Journal*, 2021 8 words — < 1%
Crossref
-
- 13 Silva, Carolina Santos, Flávia de Souza Lins Borba, Maria Fernanda Pimentel, Marcio José Coelho 6 words — < 1%

Pontes, Ricardo Saldanha Honorato, and Celio Pasquini.
"Classification of blue pen ink using infrared spectroscopy and
linear discriminant analysis", Microchemical Journal, 2012.

Crossref

14 M. Mancini, Å. Rinnan, A. Pizzi, C. Mengarelli, G. Rossini, D. Duca, G. Toscano. "Near infrared spectroscopy for the discrimination between different residues of the wood processing industry in the pellet sector", Fuel, 2018

6 words — < 1%

Crossref

EXCLUDE QUOTES ON

EXCLUDE MATCHES OFF

EXCLUDE BIBLIOGRAPHY ON