Journal of Engineering Science and Technology
Vol. 16, No. 6 (2021) 4650 - 4665

AUTHENTICATION OF JAVA PREANGER STEAMED GREEN TEA BY USING ULTRAVIOLET SPECTROSCOPY AND DISCRIMINANT ANALYSIS METHOD

DIDING SUHANDY[1,*], MEINILWITA YULIA[2]

[1]Spectroscopy Research Group (SRG), Laboratory of Bioprocess and Postharvest Engineering, Department of Agricultural Engineering, Faculty of Agriculture, The University of Lampung, Jl. Prof. Dr. Soemantri Brojonegoro No.1, Bandar Lampung, 35145, Indonesia
[2]Department of Agricultural Technology, Lampung State Polytechnic, Jl. Soekarno Hatta No. 10, Rajabasa Bandar Lampung, 35141, Indonesia
*Corresponding Author: diding.sughandy@fp.unila.ac.id

Abstract

Java Preanger steamed green tea (JPGT) is highly valuable and one of the most expensive green teas in Indonesia. In this research, the use of ultraviolet (UV) spectroscopy and discriminant analysis methods such as partial least square-discriminant analysis (PLS-DA), linear discriminant analysis (LDA) and principal components analysis-linear discriminant analysis (PCA-LDA) was evaluated for the authentication of JPGT samples. Three types of green teas were prepared for samples: 200 samples of Java Preanger steamed green tea (JPGT) with geographic indications (GIs tea), 200 samples of Chun Mee (CM) green tea grade I (non-GIs and non-JPGT tea) and 200 samples of Sow Mee (SM 315) green tea grade II (non-GIs and non-JPGT tea). The UV spectral data of all samples was acquired using a UV visible spectrometer within the 190-400 nm range. The principal component analysis (PCA) of pre-processed spectra was able to separate JPGT samples from non-JPGT samples. PLS-DA, LDA and PCA-LDA was also able to classify prediction samples with a 100% classification accuracy. This UV spectroscopy method is simpler, cheaper and acceptable for establishing JPGT authentication with potential for daily routine analysis in tea industry.

Keywords: Authentication, Java Preanger steamed green tea (JPGT), LDA, PCA-LDA, PLS-DA, UV spectroscopy.

1. Introduction

Tea is one of the oldest and most consumed beverages globally, with 6.34 million tons produced worldwide, annually [1-3]. In 2018, Asia produced 5.43 million tons of tea, of which China contributed substantially, about 41% worldwide. In general, based on the degree of fermentation processes, tea can be typically divided into three types: non-fermented (green teas), fermented (Oolong, white and black teas), and postfermented (dark teas) [4]. Green tea is one of the most enjoyable, popular and healthy beverages, rich in polyphenols especially in *Camellia sinensis var. Assamica* [5]. According to Cabrera et al. [5], about 20-22% of tea produced is green tea, and is mostly consumed in East Asia region such as China, Japan and Korea. Recently, the demand for green tea has been strongly driven by two important issues: its quality and geographical origin [6].

The Java Preanger steamed green tea (JPGT) is special-grade, and one of the most valuable and expensive green teas in Indonesia [7]. It is made from Neglasari tea plantation in the mountainous area of the West Java province, including Mount Tilu with its distinct flavour: thick astringent, brish aster alongside a tasty and steamed peanut aroma [7]. It is a high-quality product with a price ten times higher than the regular non-JPGT teas. The JPGT is highly prone to risk which includes fraud and label falsification. This is why, in 2014, the Indonesian government through Indonesian Directorate General of Intellectual Property officially registered and protected JPGT as one of Indonesian tea product with a geographic indication (GIs) [7]. Hence, it is difficult to discriminate between special-grade JPGT samples and other non-JPGT samples based on tea appearance to the naked eyes, both before and after infusion.

Traditionally, the quality and geographical origin of tea was evaluated by skilful tasters according to several parameters such as aroma, flavour and appearance including colour [8]. However, the drawbacks associated with this assessment include time-consumption, inconsistency problems and imprecise results depending highly on the conditions of the taster [8]. Consequently, it is necessary to establish an analytical method that replaces the traditional tea assessment. This method helps in developing a more precise and reliable routine analyses of tea samples in order to prevent mis-labelling.

Several analytical methods combined with various chemometrics have been proposed for the classification of teas according to its origin and grades in order to protect both producers and consumers [9] such as high-performance liquid chromatography (HPLC) [10, 11], near infrared (NIR) spectroscopy [12, 13], mid-infrared spectroscopy [14], nuclear magnetic resonance (^{1}H NMR) spectroscopy [15, 16], fluorescence spectroscopy [17, 18], electronic nose/tongue [19] and fusion methods [20]. Particularly, the discriminant analysis applied in spectral data has been used for tea authentication. Multi-wavelength statistical discriminant analysis (MW-SDA) and NIR spectroscopy was used for the identification of Shandong green tea origins with a 98.3% of accuracy [13]. Meng et al. [20] used spectral data from ^{1}H NMR and NIR spectroscopy coupled with orthogonal partial least square-discriminant analysis (OPLS-DA), partial least square-discriminant analysis (PLS-DA) and principal component analysis-linear discriminant analysis (PCA-LDA) for geographical origin discrimination of Oolong tea from three different growing places in the Fujian province of China, which resulted in an accuracy of 86.2-95.8%. The PLS-DA and NIR spectroscopy was used to develop a rapid discrimination of the geographical origins of

Anxi-Tieguanyin Oolong tea [21]. Fluorescence spectroscopy and linear discriminant analysis (LDA) was used to differentiate several Sri Lankan black teas according to their origins and elevations [22].

Compared to a large number of published works using NIR and mid-infrared spectroscopy, there was a few researches performed using UV spectroscopy for tea authentication purposes. For example, the spectral data of a methanol extract of green tea samples between 200 and 400 nm was used in association with PLS-DA to classify green teas from East and South Asia with 100% correct classification [23]. Using methanol for tea extraction, Palacios-Morillo et al. [24] utilized UV-visible spectral data within the range of 250-800 nm combined with discriminant analysis methods such as LDA to classify tea varieties (black, green and Pu-erh teas). Overall, LDA resulted in 97.7% sensitivity and 98.9% specificity, respectively [24]. Diniz et al. [25] utilized UV-visible spectroscopy of water extract of tea samples alongside several classification methods including PLS-DA, PCA-LDA and successive projection algorithm associated with linear discriminant analysis (SPA-LDA) for geographical and varietal classification of black and green teas with acceptable results. The raw spectral data within the 251-490 nm range resulted in a better classification rate compared to full spectrum (190-800 nm). UV-visible spectroscopy is advantageous compared to other spectroscopic methods. It is a green technology without chemical waste (applicable in water extraction), inexpensive spectroscopy and relatively fast spectral measurement [26, 27].

In this study, the spectral data in UV region of water extract tea samples coupled with three discriminant analysis methods of PLS-DA, LDA and PCA-LDA was applied to differentiate between JPGT and non-JPGT green tea samples. Furthermore, the aim was to acquire spectral data of JPGT and non-JPGT green tea samples within the 200-400 nm range in order to obtain a faster spectral acquisition process. Several spectral pre-processing methods were also applied to develop a better and robust classification model.

2. Materials and Methods

2.1. Samples and extraction protocol

Six hundred green tea samples (one gram each) were prepared from two distinct tea plantations and two different tea processing companies in West Java province, Indonesia as seen in Fig. 1 and Table 1. In order to avoid the spectral variation due to different tea harvesting time, all samples were collected from the same harvest season in 2019. To show ability of UV spectroscopy in the authentication of JPGT, it is important to use non-JPGT tea samples with a wide range of various green tea grades. The samples have three different grade levels: 200 samples belong to Java Preanger steamed green tea (JPGT) with geographic indications (GIs tea), 200 samples belong to Chun Mee (CM) green tea grade I (non-JPGT tea) and 200 samples belong to Sow Mee (SM) 315 green tea grade II (non-JPGT tea). They were finely grounded using home grinder (Sayota) and sieved through a nest of U. S. standard sieves (mesh number of 50) on a Meinzer II sieve shaker (CSC Scientific Company, Inc. USA) for 10 minutes to obtain the uniform size of tea particle of 297 micrometer. The sample features are described in Table 1. The protocol for extraction of tea samples was done using a hot distilled water (water temperature at 90-98°C) according to Diniz et al. [25] with a minor modification. One gram of tea sample was prepared in 100 mL of distilled water at 90-98°C and stirred using a magnetic stirrer (CiblancTM Magnetic Stirrer) for 10 minutes at 1750 rpm. Tea extraction was filtered and diluted with distilled water in a proportion of 1:80 (tea extraction: distilled water). Diluted samples are ready for spectral analysis. A complete tea extraction protocol used in this research was depicted in Fig. 2.

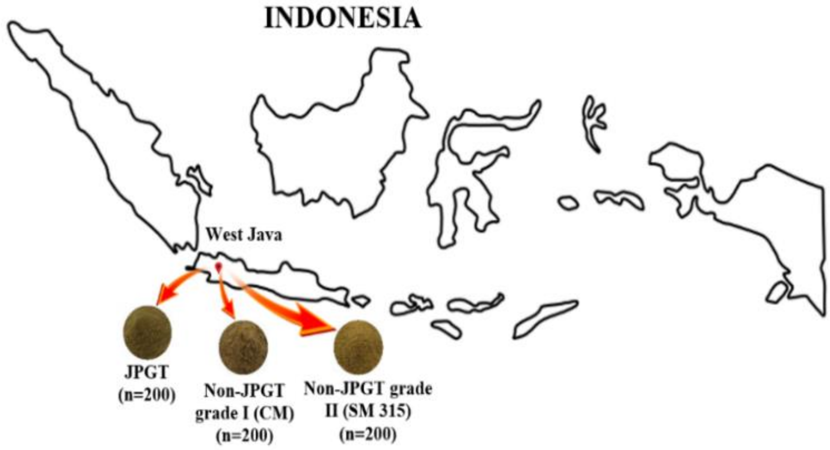

Fig. 1. Map of West Java, Indonesia showing the geographical origin of tea samples used in this research.

Table 1. The characteristics of tea samples used in this research.

	Java Preanger tea	**Chun Mee (CM) tea**	**Sow Mee (SM) 315 tea**
Number of samples	200	200	200
Variety	*Camellia sinensis var. Assamica*	*Camellia sinensis var. Assamica*	*Camellia sinensis var. Assamica*
Type of tea	Steamed green tea	Steamed green tea	Steamed green tea
GIs (Geographic Indications)	GIs	Non-GIs	Non-GIs
Grade	Grade 1	Grade 1	Grade II
Mesh	50	50	50
Plantation name	Neglasari, West Java, Indonesia	Dewata, West Java, Indonesia	Dewata, West Java, Indonesia
Longitude	107°31’57.032”	107°28'41.7"	107°28'41.7"
Latitude	07°10’55.623”	07°12'43.4"	07°12'43.4"
Altitude (masl)	1498-1520	1350–1500	1350–1500
Soil type	Andisol	Andisol	Andisol
Tea processing company	Gambung Agro Lestari	Kabepe Chakra	Kabepe Chakra
Appearance			

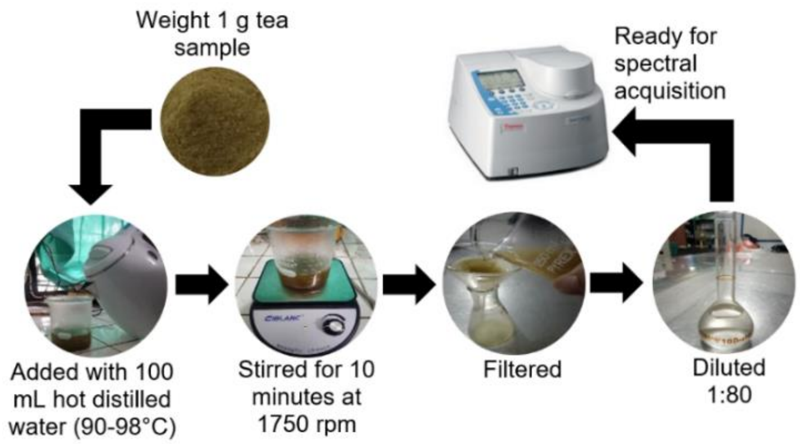

Fig. 2. The tea extraction protocol using hot distilled water according to Diniz et al. [25] with a minor modification.

2.2. UV spectral data collection

Extracted and diluted tea samples (tea infusions, 3 mL) were used for spectral measurement. The UV spectral data of 600 tea samples within the 190-400 nm range and 1 nm of resolution were collected in transmission mode using a dual-beam UV-Visible benchtop spectrometer (Genesys 10s UV-Vis, Thermo Scientific Inc., Madison, WI) supplied with a quartz cell with a 10 mm optical path. The spectrometer had a high-intensity xenon lamp as light source and dual Silicon photodiodes as a detector. The spectral measurement was done in room temperature conditions of about 28°C. The absorbance values (A) were determined from transmittance data (T) using the following formula [28]:

$$A = \log\frac{1}{T} \tag{1}$$

2.3. Discriminant analysis methods

The obtained UV spectral data from 600 samples over the wavelength range of 190-400 nm were subjected to principal component analysis (PCA) through an unsupervised pattern recognition method. The scores and *x*-loading plots were used to investigate possible sample differences between JPGT and non-JPGT samples. To evaluate the differences among tea samples, three supervised classification multivariate statistical methods belonging to the discriminant analysis-based methods were used namely, PLS-DA, LDA and PCA-LDA. Detailed explanation on those methods is found in several past research [29-31].

To apply discrimination methods, the samples were split randomly into two subsets: calibration and validation (n=501) and prediction (n=99). For PLS-DA, each sample was labelled with an arbitrary number indicating the sample category, either JPGT or non-JPGT samples (JPGT=1 and non-JPGT=0). Therefore, tea samples belong to either JPGT or non-JPGT depending on its value between 0.5-1.5 and 0.5-(-0.5), respectively. For LDA, several wavelengths were selected from PCA *x*-loading plot as input in *x* variables (predictor). In order to develop a robust classification model, the PCA and discriminant analysis methods were performed on both raw and pre-processed spectra. Three pre-processing algorithms were used namely, multiplicative scatter correction (MSC), 9 points moving averaging smoothing (MAS 9s) and combination of MSC and MAS 9s (MSC+MAS 9s). The MSC is an effective spectral pre-processing algorithm which compensates for additive and multiplicative effects [32]. The MAV 9s was used to increase signal noise to the ratio (SNR) of the obtained UV spectral data [32]. Every multivariate statistical analysis conducted, including spectral pre-processing, PCA and discriminant methods were performed using The Unscrambler X version 10.4 (64-bit) (Camo Software AS, Oslo, Norway). The performance of discriminant analysis methods was evaluated by calculating the accuracy rate for each model both in raw and pre-processed spectral data. The accuracy rate (%) is mathematically expressed below [33]:

$$\text{Accuracy rate (\%)} = \frac{\text{Number of correctly classified samples}}{\text{Total number of samples}} \times 100\% \qquad (2)$$

3. Results and Discussion

3.1. Spectral analysis

Figure 3 demonstrates the raw spectra of samples with the different types of green teas. The spectral data were similar to that of previous reported work both in shape and intensity [24-25, 34]. Due to this similarity, it is difficult to differentiate between JPGT and non-JPGT directly. Within the window of 190-220 nm the spectral data was highly noisy and therefore removed from further analysis. Consequently, the window at 220-400 nm was selected for further analysis. This spectral window is closely related to the n→π electronic transition of methylxanthines and catechins compounds in tea samples as reported by previous research [24]. In the spectral window, two peaks were observed within the wavelengths of 250 nm with approximately 0.3 of absorbance intensity, and 272 nm with 0.7 absorbance intensity. The peaks at 250 and 272 nm correspond to phenolic compounds such as gallic, chlorogenic and caffeic acids present in the green tea samples [23, 25].

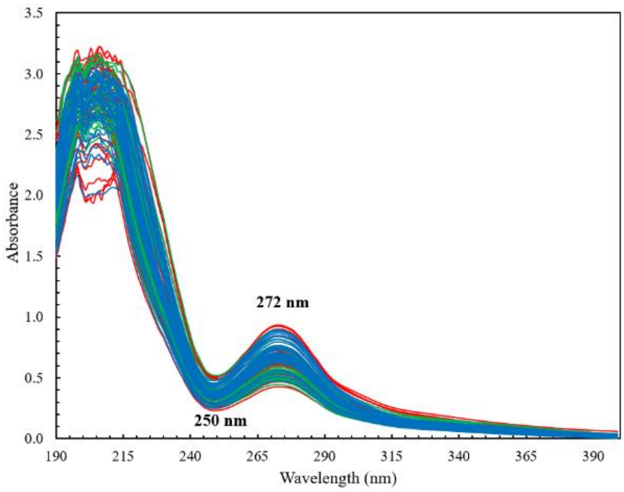

Fig. 3. The raw spectra of JPGT (in red line), non-JPGT grade I (in blue line) and non-JPGT grade II (in green line) within the 190-400 nm range.

Figure 4 shows the pre-processed spectra (MSC+MAV 9s) of the different green tea types within the range of 220-400 nm. A clear difference between JPGT and non-JPGT tea samples was observed in the 272 nm wavelength which corresponds with the difference in phenolic contents [23, 25].

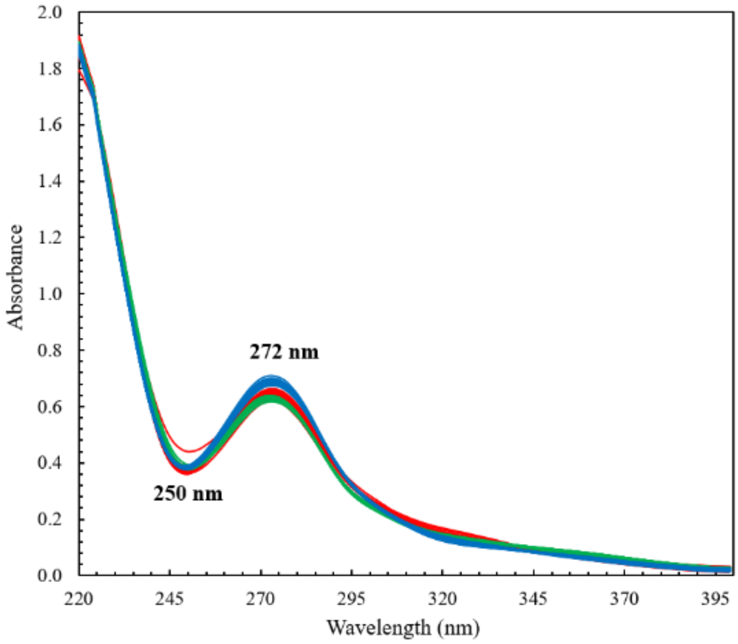

Fig. 4. The pre-processed spectra of JPGT (in red line), non-JPGT grade I (in blue line) and non-JPGT grade II (in green line) within the 220-400 nm range.

3.2. The PCA results

Figure 5 shows the PCA score plots of PC1, PC2 and PC3 using raw UV spectral data (220-400 nm) for 600 samples, which were calculated using nonlinear iterative partial least squares (NIPALS) algorithm with a full-cross validation method [35]. Consider a data matrix X of absorbance data (*A*) with *n* columns (samples) and *m* rows (variables). In present study, *n*= 600 (200 samples for each tea types) and *m*=180 (180 variables, ranging from 220 nm to 400 nm, in steps of 1 nm). PCA is mathematically defined as an orthogonal linear transformation that transforms the *X* data to a new coordinate system such that the greatest variance by some projection of the data lies on the first coordinate (first principal component—PC1), the second greatest variance lies on the second coordinate (second principal component—PC2), and so on. The mathematical model corresponding to PCA is based on the decomposition of the X matrix into the $n \times A$ score matrix (T) and $m \times A$ loading matrix (P) as Eq. (3) [36]:

$$X = TP' + F = \sum_{a=1}^{A} t_a p'_a + F \quad (3)$$

where *X* is the spectral data matrix, *T, P* and *F* is known as the scores, the *x*-loadings and the residual matrix, respectively. t_a is the sample score vector on each principal component (PC) for *X*, and p_a is the vector of the variable loading on each PC for *X*. It can be said that the loadings matrix P is the weights or influence of the variables in matrix X on the matrix scores T. The *x*-loadings plot could be called a map of variables while the scores plot is a map of samples.

In Fig. 5, it was clear that no significant separation was observed among JPGT and non-JPGT green tea samples, although the accumulated variance of the first three PCs was 99% (PC1, PC2 and PC3 explained 97, 1 and 1% of the total variance, respectively). PCA was also applied on the UV spectral data of all samples (n=600) within the 220-400 nm range after multiplicative scatter correction and moving average smoothing in 9 segments of the raw data.

The result was plotted in Fig. 6 showing the scores of the first three principal components (PC1xPC2xPC3). The PCA successfully discriminated the green tea samples into three well defined clusters with PC1=71%, PC2=16% and PC3=8% of the data variance, respectively. The JPGT green tea sample cluster was located in the middle part between PC1 positive and PC1 negative. The normal tea samples grade I (non-JPGT) were found in the negative PC1 (PC1<0), and normal tea samples grade II (non-JPGT) were positioned at the positive PC1 (PC1>0).

A similar result was reported by Aboulwafa et al. [23] where the spectral data between 200 and 400 nm alongside the PCA successfully differentiated between South and East Asian green tea samples using PC1 and PC2 with 94% of the total data variance [23]. Thus, it was proven that the pre-processed spectral data gives better results for classification between JPGT and non-JPGT samples.

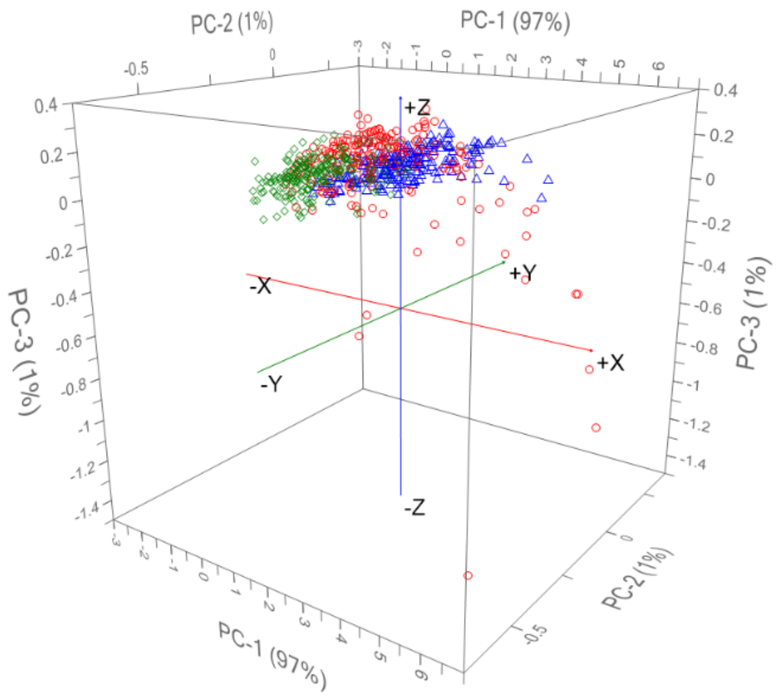

Fig. 5. The PCA scores plot of PC1, PC2 and PC3 using raw UV spectral data (220-400 nm) of JPGT (red circle), non-JPGT grade I (blue open triangle) and non-JPGT grade II (green open diamond).

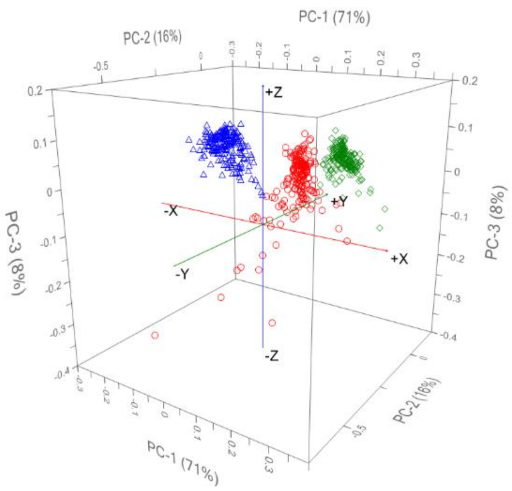

Fig. 6. The PCA scores plot of PC1, PC2 and PC3 using preprocessed (MSC+MAV 9s) UV spectral data (220-400 nm) of JPGT (red circle), non-JPGT grade I (blue open triangle) and non-JPGT grade II (green open diamond).

Figure 7 shows the *x*-loadings for the first three principal components (PCs) calculated using pre-processed spectra showing the contribution of each wavelength (220-400 nm) in the differentiation between JPGT and non-JPGT samples. In general, it is observed that the spectral window at 220-300 nm with higher *x*-loadings showed a stronger ability to discriminate tea samples compared to the window at 300-400 nm. Moreover, there were several peaks with high *x*-loadings. The peaks with positive loadings were observed at 235 nm, 255 nm, 300 nm and 360 nm. The peaks with negative loadings were identified at 230 nm, 242 nm and 275 nm. These peaks contributed greatly in differentiating between JPGT and non-JPGT samples and were used in LDA calculation. Previously, Aboulwafa et al. [23] discovered a maximum absorbance at 273 nm and 330 nm which is due to the phenolic acids in green tea samples.

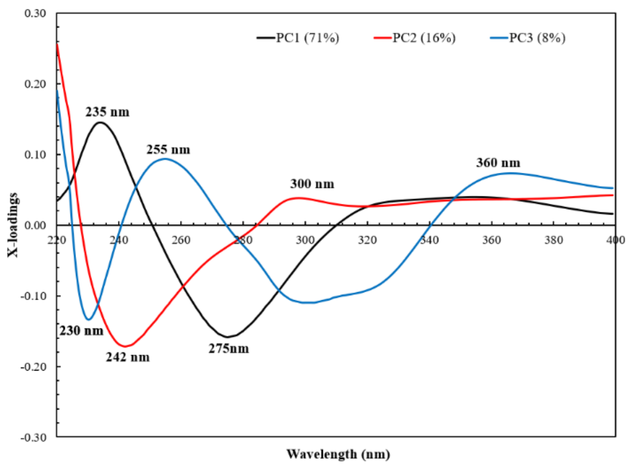

Fig. 7. The *x*-loadings plot for the first three PCs calculated using pre-processed spectra (MSC+MAV 9s) within the 220-400 nm range.

3.3. The classification results

Three classifier models based on the discriminant analysis methods, namely PLS-DA, LDA and PCA-LDA were used in this research to classify JPGT and non-JPGT samples. The calculation in each method was performed on raw and pre-processed spectra as seen in Table 2, which was for calibration and Table 3 for prediction, respectively. In developing the PLS-DA model, the optimal number of latent variables (LVs) was selected using the lowest root mean square error of validation (RMSEV). For this, the LVs for PLS-DA models were 10, 12, 9, and 10 for raw, MSC, MAS 9s and MSC+MAS 9s spectral data, respectively. All PLS-DA models obtained a classification accuracy of 99.80% for raw and pre-processed spectral data. There was only one sample of JPGT erroneously classified as a non-JPGT sample.

Table 2. Classification performance using discriminant analysis methods of PLS-DA, LDA and PCA-LDA with raw and preprocessed spectra within the 220-400 nm range for calibration and validation set (n=501).

Model	Pre-processing	Correctly classified sample JPGT (n=167)	Non-JPGT (n=334)	Accuracy rate (%)
PLS-DA	RAW	166	334	99.80
PLS-DA	MAS	166	334	99.80
PLS-DA	MSC	166	334	99.80
PLS-DA	MSC+MAS	166	334	99.80
LDA	RAW	167	332	99.60
LDA	MAS	166	331	99.20
LDA	MSC	164	333	99.20
LDA	MSC+MAS	162	333	98.80
PCA-LDA	RAW	167	334	100.00
PCA-LDA	MAS	167	334	100.00
PCA-LDA	MSC	167	334	100.00
PCA-LDA	MSC+MAS	167	334	100.00

Table 3. Classification performance using discriminant analysis methods of PLS-DA, LDA and PCA-LDA with raw and preprocessed spectra within the 220-400 nm range for prediction set (n=99).

Model	Pre-processing	Correctly classified sample JPGT (n=33)	Non-JPGT (n=66)	Accuracy rate (%)
PLS-DA	RAW	33	66	100.00
PLS-DA	MAS	33	66	100.00
PLS-DA	MSC	33	66	100.00
PLS-DA	MSC+MAS	33	66	100.00
LDA	RAW	33	66	100.00
LDA	MAS	33	66	100.00
LDA	MSC	33	66	100.00
LDA	MSC+MAS	33	66	100.00
PCA-LDA	RAW	33	66	100.00
PCA-LDA	MAS	33	66	100.00
PCA-LDA	MSC	33	66	100.00
PCA-LDA	MSC+MAS	33	66	100.00

LDA models for raw and pre-processed spectral data was performed using seven important wavelengths obtained from PCA *x*-loadings as input: 230 nm, 235 nm, 242 nm, 255 nm, 275 nm, 300 nm and 360 nm. For LDA, the optimal model was achieved using raw spectral data with a 99.60% accuracy rate. In this model, two samples of non-JPGT were erroneously classified as JPGT samples.

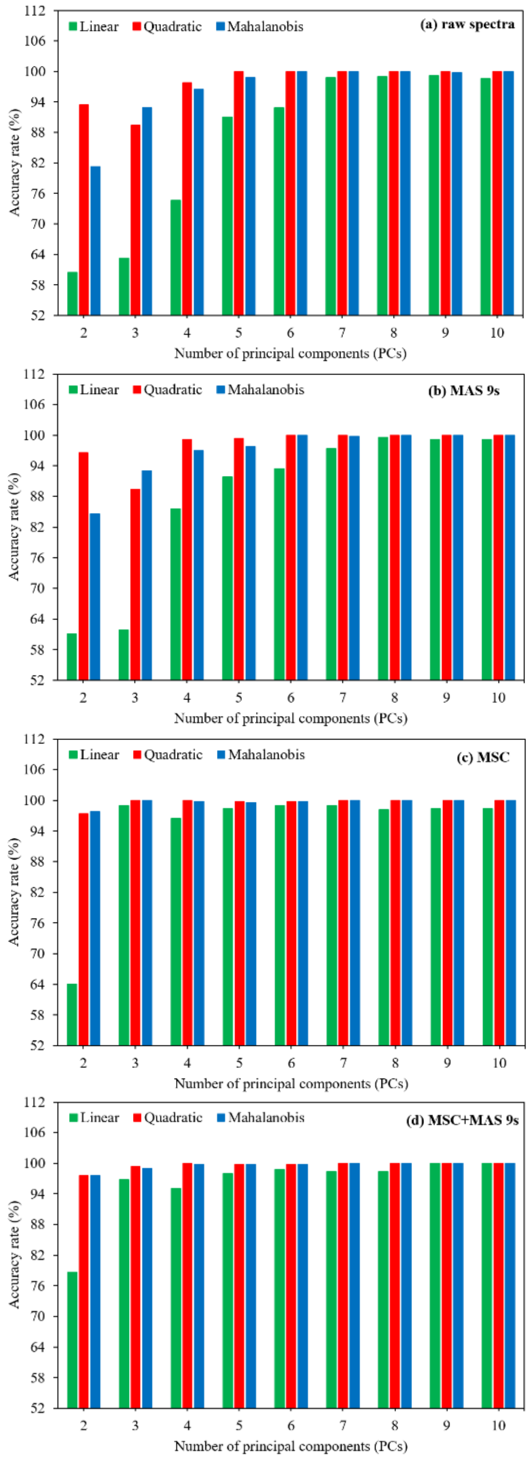

Fig. 8. The influence of number of PCs and calculation method of PCA-LDA on accuracy rate for raw and pre-processed spectra.

In developing the PCA-LDA model, two parameters have to be optimized: the number of principal components (PCs) and method of distance calculation (linear, quadratic and Mahalanobis). As seen in Fig. 8, the accuracy of PCA-LDA was highly influenced by both parameters. For raw spectra, the optimal parameter was the quadratic method with five PCs which resulted in a 100% accuracy (Fig. 8). For MAS 9s spectra, the optimal parameter was obtained using two conditions: quadratic method with six PCs and Mahalanobis method with six PCs. Both conditions resulted in a 100% accuracy as seen in Fig. 8. A similar result was observed in the MSC spectra. Two optimal conditions which resulted in 100% accuracy were achieved using quadratic method with three PCs and Mahalanobis method with three PCs. The optimal parameter for the MSC+MAS 9s spectra was the quadratic method with four PCs which resulted in a 100% accuracy. Overall, the best PCA-LDA model was achieved for MSC spectra using quadratic or Mahalanobis method with three PCs. There were no erroneously classified samples, which resulted in the 100% accuracy rate. From this result, it is therefore concluded that all discriminant analysis methods are powerful and effective means of classifying JPGT and non-JPGT samples. Overall, all developed models were acceptable with 100% of accuracy rate in prediction as seen in Table 3.

The accuracy rate obtained in our research agrees with previously published works. An accuracy rate of 100% was obtained for classifying green tea samples from East and South Asia using UV spectral data between 200 and 400 nm combined with PLS-DA [23]. Using a wavelength of 251-490 nm, a classification of black and green tea samples from different origins was achieved with a 68% accuracy rate for PLS-DA, and 100% for PCA-LDA [25]. A PLS-DA model of two different grades of black CTC tea can be developed using pre-processed spectral data within the 190-1100 nm range which resulted in 100% prediction accuracy [34]. Ni et al. [37] utilized combined multi-element and isotopic profile to discriminate Xihu Longjing green tea from other green teas. PLS-DA and LDA were used for developing classification models which resulted in an accuracy rate of 82% and 89%, respectively [37]. Recently, an element analyser isotope ratio mass spectrometry (EA-IRMS) and inductively coupled plasma mass spectrometry (ICP-MS) were used along with chemometrics to authenticate Chongqing tea with different ingredients. The best model was obtained by LDA, and the total authentication rates for model training, cross-validation and external validation were 100%, 100% and 96.9%, respectively [38]. Using similar methods, Liu et al. [39] developed classification models including PLS-DA and LDA to authenticate Pu'er tea in different production years. The best performance was achieved using LDA model with 100% of accuracy rate. Xu et al. [40] employed LDA on the e-nose data of several volatile components of tea leaves and infusions to discriminate six different grades of Longjing tea, which resulted in 100% accuracy.

4. Conclusions

Using an unsupervised method of PCA, the clustering ability of JPGT and non-JPGT samples using pre-processed spectral data within the 220-400 nm range was better than that of the raw spectral data. This study therefore demonstrated the potential use of UV spectroscopy combined with three discriminant analysis methods of PLS-DA, LDA and PCA-LDA to differentiate between JPGT and non-JPGT samples. The accuracy rate of all classification models was high both in calibration and prediction set. In conclusion, a simple, relatively fast and affordable analytical method has been developed based on UV spectroscopy and discriminant analysis methods to distinguish JPGT samples from non-JPGT in order to protect JPGT from any fraud and mis-labelling. This procedure may replace the traditional method of tea evaluation with a potential for daily routine analysis in the tea industry.

Acknowledgments

The authors express their gratitude to the Kemdikbudristek, the Republic of Indonesia, for their financial support. The authors are also grateful to PT. Kabepe Chakra for the tea samples and information.

Nomenclatures

A	Absorbance
T	Transmittance

Abbreviations

CM	Chun Mee
EA-IRMS	Element Analyzer Isotope Ratio Mass Spectrometry
GIs	Geographic Indication
HPLC	High-Performance Liquid Chromatography
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
JPGT	Java Preanger Steamed Green Tea
LDA	Linear Discriminant Analysis
LV	Latent Variable
MAS	Moving Averaging Smoothing
MSC	Multiplicative Scatter Correction
MW-SDA	Multi-Wavelength Statistical Discriminant Analysis
NIPALS	Nonlinear Iterative Partial Least Squares
NIR	Near Infrared
NMR	Nuclear Magnetic Resonance
OPLS-DA	Orthogonal Partial Least Square-Discriminant Analysis
PC	Principal Component
PCA	Principal Component Analysis
PCA-LDA	Principal Components Analysis-Linear Discriminant Analysis
PLS-DA	Partial Least Square-Discriminant Analysis
RMSEV	Root Mean Square Error of Validation
SM	Sow Mee
SNR	Signal-to-Noise Ratio
UV	Ultraviolet
UV-VIS	Ultraviolet-Visible

References

1. Lin, X.; and Sun, D. (2020). Recent developments in vibrational spectroscopic techniques for tea quality and safety analyses. *Trends in Food Science and Technology*, 104, 163-176.
2. FAO. (2020). *Food and agriculture data*. Retrieved August 20, 2020, from http://www.fao.org/faostat/en/#data/QC.

3. Pillay, A.E.; and Stephen, S. (2018). Application of hyphenated plasma mass spectrometry to the study of elevated metal toxins in brewed beverages. *Journal of Engineering Science and Technology*, 13(10), 3359-3368.
4. Engelhardt, U.H. (2020). Tea chemistry – What do and what don't we know? A micro review. *Food Research International*, 132, 109120.
5. Cabrera, C.; Artacho, R.; and Giménez, R. (2006). Beneficial effects of green tea – A review. *Journal of the American College of Nutrition*, 25(2), 79-99.
6. Li, C.; Guo, H.; Zong, B.; He, P.; Fan, F.; and Gong, S. (2019). Rapid and non-destructive discrimination of special-grade flat green tea using near-infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 206, 254-262.
7. DGIP. (2020). *Geographic indications*. Retrieved August 20, 2020, from http://e-book.dgip.go.id/indikasi-geografis/?book=mpig-teh-java-prearger.
8. Diniz, P.H.G.D.; Gomes, A.A.; Pistonesi, M.F.; Band, B.S.F.; and De Araújo, M.C.U. (2014). Simultaneous classification of teas according to their varieties and geographical origins by using NIR spectroscopy and SPA-LDA. *Food Analytical Methods*, 7(8), 1712-1718.
9. Ye, N.S. (2012). A minireview of analytical methods for the geographical origin analysis of teas (Camellia sinensis). *Critical Reviews in Food Science and Nutrition*, 52(9), 775-780.
10. Song, M.; Li, Q.; Guan, X.; Wang, T.; and Bi, K. (2013). Novel HPLC method to evaluate the quality and identify the origins of Longjing green tea. *Analytical Letters*, 46(1), 60-73.
11. Fang, S.; Huang, W.; Wei, Y.; Tao, M.; Hu, X.; Li, T.; Kalkhajeh, Y.K.; Deng, W.; and Ning, J. (2019). Geographical origin traceability of Keemun black tea based on its non‐volatile composition combined with chemometrics. *Journal of the Science of Food and Agriculture*, 99(15), 6937-6943.
12. Firmani, P.; De Luca, S.; Bucci, R.; Marini, F.; and Biancolillo, A. (2019). Near infrared (NIR) spectroscopy-based classification for the authentication of Darjeeling black tea. *Food Control*, 100, 292-299.
13. Zhuang, X.G.; Shi, X.S.; Wang, H.F.; Wang, L.L.; and Fang, J.X. (2019). Rapid determination of green tea origins by near-infrared spectroscopy and multi-wavelength statistical discriminant analysis. *Journal of Applied Spectroscopy*, 86, 76-82.
14. Wei, L.; Hu, O.; Chen, H.; Yang, T.; Fan, Y.; Xu, L.; Zhang, L.; Lan, W.; She, Y.; and Fu, H. (2020). Variety identification and age prediction of Pu-erh tea using graphene oxide and porphyrin complex based mid-infrared spectroscopy coupled with chemometrics. *Microchemical Journal*, 158, 105255.
15. Le Gall, G.; Colquhoun, I.J.; and Defernez, M. (2004). Metabolite profiling using ^{1}H NMR spectroscopy for quality assessment of green tea, Camellia sinensis (L.). *Journal of Agricultural and Food Chemistry*, 52(4), 692-700.
16. Lee, J.E.; Lee, B.; Chung, J.; Hwang, J..; Lee, S.J.; Lee, C.; and Hong, Y. (2010). Geographical and climatic dependencies of green tea (*Camellia sinensis*) metabolites: A ^{1}H NMR-based metabolomics study. *Journal of Agricultural and Food Chemistry*, 58(19), 10582-10589.
17. Zhu, J.; Zhu, F.; Li, L.; Cheng, L.; Zhang, L.; Sun, Y.; Wan, X.; and Zhang, Z. (2019). Highly discriminant rate of Dianhong black tea grades based on fluorescent probes combined with chemometric methods. *Food Chemistry*, 298, 125046.

18. Hooshyari, M.; Rubio, L.; Casale, M.; Furlanetto, S.; Turrini, F.; Sarabia, L.A.; and Ortiz, M.C. (2019). D-optimal design and PARAFAC as useful tools for the optimisation of signals from fluorescence spectroscopy prior to the characterisation of green tea samples. *Food Analytical Methods*, 12, 761-772.
19. Banerjee, M.B.; Roy, R.B.; Tudu, B.; Bandyopadhyay, R.; and Bhattacharyya, N. (2018). Black tea classification employing feature fusion of e-nose and e-tongue responses. *Journal of Food Engineering*, 244, 55-63.
20. Meng, W.; Xu, X.; Cheng, K.; Xu, J.; Shen, G.; Wu, Z.; and Dong, J. (2017). Geographical origin discrimination of oolong tea (TieGuanYin, *Camellia sinensis* (L.) O. Kuntze) using proton nuclear magnetic resonance spectroscopy and near-infrared spectroscopy. *Food Analytical Methods*, 10(11), 3508-3522.
21. Yan, S.M.; Liu, J.P.; Xu, L.; Fu, X.S.; Cui, H.F.; Yun, Z.Y.; Yu, X.P.; and Ye, Z.H. (2014). Rapid discrimination of the geographical origins of an Oolong tea (Anxi-Tieguanyin) by near infrared spectroscopy and partial least squares discriminant analysis. *Journal of Analytical Methods in Chemistry*, 2014, 704971.
22. Seetohul, L.N.; Scott, S.M.; O'Hare, W.T.; Ali, Z.; and Islam, M. (2013). Discrimination of Sri Lankan black teas using fluorescence spectroscopy and linear discriminant analysis. *Journal of the Science of Food and Agriculture*, 93(9), 2308-2314.
23. Aboulwafa, M.M.; Youssef, F.S.; Gad, H.A.; Sarker, S.D.; Nahar, L.; Al-Azizi, M.M.; and Ashour, M.L. (2019). Authentication and discrimination of green tea samples using UV-visible, FTIR and HPLC techniques coupled with chemometrics analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 164, 653-658.
24. Palacios-Morillo, A.; Alcázar, Á.; De Pablos, F.; and Jurado, J.M. (2013). Differentiation of tea varieties using UV–vis spectra and pattern recognition techniques. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 103, 79-83.
25. Diniz, P.H.G.D.; Barbosa, M.F.; De Melo Milanez, K.D.T.; Pistonesi, M.F.; and De Araújo, M.C.U. (2016). Using UV–vis spectroscopy for simultaneous geographical and varietal classification of tea infusions simulating a home-made tea cup. *Food Chemistry*, 192, 374-379.
26. Suhandy, D.; and Yulia, M. (2021). The use of ultraviolet (UV) spectroscopy and chemometrics to quantify the percentages of adulteration in Kalosi ground roasted specialty coffee. *Journal of Engineering Science and Technology*, 16(1), 350-364.
27. Suhandy, D.; and Yulia, M. (2017). Peaberry coffee discrimination using UV-visible spectroscopy combined with SIMCA and PLS-DA. *International Journal of Food Properties*, 20(sup1), S331-S339.
28. Huang, Y.; Dong, W.; Sanaeifar, A.; Wang, X.; Luo, W.; Zhan, B.; Liu, X.; Li, R.; Zhang, H.; and Li, X. (2020). Development of simple identification models for four main catechins and caffeine in fresh green tea leaf based on visible and near-infrared spectroscopy. *Computers and Electronics in Agriculture*, 173, 105388.
29. Brereton, R.G. (2009). *Chemometrics for pattern recognition.* West Sussex: Wiley.
30. Oliveri, P.; and Downey, G. (2013). Discriminant and class-modelling chemometric techniques for food PDO verification. *Comprehensive Analytical Chemistry*, 60, 317-338.

31. Barker, M.; and Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, 17(3), 166-173.
32. Verboven, S.; Hubert, M.; and Goos, P. (2012). Robust preprocessing and model selection for spectral data. *Journal of Chemometrics*, 26(6), 282-289.
33. Li, M.; Pan, T.; and Chen, Q. (2020). Estimation of tea quality grade using statistical identification of key variables. *Food Control*, 119, 107485.
34. Suhandy, D.; and Yulia, M. (2019). Potential application of UV-visible spectroscopy and PLS-DA method to discriminate Indonesian CTC black tea according to grade levels. *IOP Conference Series: Earth and Environmental Science*, 258, 012042.
35. Wold, S.; Geladi, P.; Esbensen, K.; and Ohman, J. (1987). Multi-way principal components-and PLS-analysis. *Journal of Chemometrics*, 1(1), 41-56.
36. Nazarloo, A.S.; Sharabiani, V.R.; Gilandeh, Y.A.; Taghinezhad, E.; Szymanek, M.; and Sprawka, M. (2021). Feasibility of using vis/nir spectroscopy and multivariate analysis for pesticide residue detection in tomatoes. *Processes*, 9(196), 1-13.
37. Ni, K.; Wang, J.; Zhang, Q.; Yi, X.; Ma, L.; Shi, Y.; and Ruan, J. (2018). Multi-element composition and isotopic signatures for the geographical origin discrimination of green tea in China: A case study of Xihu Longjing. *Journal of Food Composition and Analysis*, 67, 104-109.
38. Liu, H.; Zeng, Y.; Yan, J.; Huang, R.; Zhao, X.; Zheng, X.; Mo, M.; Tan, S.; and Tong, H. (2020). C N H O and mineral element stable isotope ratio analysis for authentication in tea. *Journal of Food Composition and Analysis*, 91, 103513.
39. Liu, H.; Zeng, Y.; Zhao, X.; and Tong, H. (2020). Chemometric authentication of Pu'er teas in terms of multielement stable isotope ratios analysis by EA-IRMS and ICP-MS. *Journal of Food and Drug Analysis*, 28(2), 248-260.
40. Xu, M.; Wang, J.; and Zhu, L. (2021). Tea quality evaluation by applying e-nose combined with chemometrics methods. *Journal of Food Science and Technology*, 58, 1549-1561.