Indonesian Black Tea Classification Using Fourier-Transform Near-Infrared Spectroscopy and a Principal Component Analysis

R O Anindya¹, J Muninggar⁴, and F S Rondonuwu²*  
¹Department of Science and Mathematics, Universitas Kristen Satya Wacana, Jl. Diponegoro 52-60, Salatiga 50711, Indonesia  
²Research Center of Near-Infrared Applications, Universitas Kristen Satya Wacana, Jl. Diponegoro 52-60, Salatiga 50711, Indonesia  
*Corresponding author’s email: ferdy@staff.uksw.edu

Abstract. Indonesian black tea (Camellia sinensis) is processed based on several standard methods to achieve a certain grade. There are more than 25 grades produced including Dust one (D1), Fanning's (FANN) and Pekoe Fanning's (PF). In general, the physical appearances are similar to each other, which makes it difficult to be differentiated from visual observations by non-experts. During the packaging process, a tea grade can easily be packed using a bag with a different label of the grade. In the global market, such a mistake should be avoided before shipping. Since shipping usually involves a large number of bags, then rapid, accurate, and nondestructive methods are critical. This paper discusses a method for classifying tea grades rapidly, accurately and nondestructively using Fourier-transform near-infrared spectroscopy (FT-NIRS) and a principal component analysis (PCA). Sixty-nine tea samples from grade Dust one (D1), Fanning's (FANN) and Pekoe Fanning's (PF) were separated into two groups; 46 samples were used for the training sets and 30 samples for the validation sets. The analysis successfully differentiated that of 3 different grades, indicating that the method is a potential alternative for tea classification in quality control processes.

Keywords: Indonesian black tea, FTIR, principal, quality control

1. Introduction
By 2015 Indonesia was the seventh largest exporter of tea in the world after Sri Lanka, China, India, Malawi, and Vietnam. It was a rank lower than that in 2012 where Indonesia occupied the sixth position. The decline in the export quantity is due to the decline in the quality of black tea (Camellia sinensis) [1]. The quality of black tea is determined by several characteristics: particle size, appearance, color, liquor taste and appearance in the infusion. The two main factors affecting the grading of black tea are the processing methods (orthodox, and CTC) and the tea particle sizes [2]. Tea processing is done through several stages, namely: the acceptance of raw materials, sifting, grinding, fermentation, drying, and sorting [3-5].

The tea processing in the mass production process with long stages allows the occurrence of errors, such as a mismatch of the content with the grade labels on the packaging, resulting in an inconsistent product quality that decreases the level of market appreciation, which further reduces sales prices.
Each type of tea has a different quality and different selling price; therefore to reduce errors in the production line, a fast, as possible, accurate, and non-destructive examination is absolutely required. Thus, a method using Fourier-transform near-infrared (FT-NIR) spectroscopy and PCA (principal component analysis) to a set of black tea samples were developed.

FT-NIR spectroscopy is well known and widely applied in specific areas of analysis, such as the analysis and identification of components in pharmaceuticals, food, and beverages [6-8]. Those applications are possible because of the FT-NIR ability to penetrate the solid and liquid samples with an electromagnetic spectrum that reaches around 1000 - 2.500 nm or an energy equivalent of about 10.000 - 4.000 cm\(^{-1}\). Research using FT-NIR was also applied in the food study that examined the ratio of sugar content in mango Gedong Gincu [9]. In health, FT-NIR was used to measure uric acid levels in urine using enzymatic methods [10]. With its wide capabilities and applications, FT-NIR was utilized to examine dry and solid tea samples.

This paper reports study of a rapid and non-destructive technique for classifying black tea using FT-NIR spectroscopy and PCA using three different black tea grades from Dust one (D1), Fanning’s (FANN), and Pekoe Fanning’s one (PFI).

2. Experimental Methods
The method used in this study was basically similar to the one developed by Shou-He Yan [11] with modifications. Three types of tea from various estates in Indonesia namely D1, FANN, and PFI were used as a set of samples, consisting of 96 data sets divided into two groups of 66 training sets in which each grade in training sets consists of 22 samples and 30 validation sets in which each grade in validation sets consists of 10 samples, vide infra. Each of the tea samples was stored in an aluminum package to avoid light radiations. Before performing the measurements, each tea sample was placed into a small container and mixed for sample homogeneity. After the mixing process, the tea samples (2-3 grams each) were put into a measuring tube (vial).

Spectral acquisitions were performed using a near infrared spectrometer Buchi Nirflx 500 solid® that operated on a 4.000-10.000 cm\(^{-1}\) energy span. The samples were loaded into vials that were ready for spectral acquisition with the utilization of NIRware spectrometer control software. Prior to analysis using the PCA method, the spectra were initially smoothened by the Savitszky-Golay method that was applied to the 1\(^{st}\) and the 2\(^{nd}\) derivative signals. The aim of the 1\(^{st}\) derivation of the spectrum was to correct the signal baseline by scattering and multiplication, while the 2\(^{nd}\) derivation aimed to increase the separation of each absorption peak. After the smoothing of the spectrum, the spectral data were analyzed using PCA. The method of PCA has been clearly explained elsewhere [12]. The results of the PCA analysis were then displayed in three-dimensional space by PC1, PC2, and PC3 axis.

3. Results and Discussion
A total of 96 tea samples, as explained in the previous section, consisted of grade D1, FANN, and PFI. From each grade, data were retrieved from 22 training sets (TS) and 10 validation sets (VS), as listed in Table 1. The TS was served as a reference for classification, whereas the VS was used to confirm the results of the classification based on the analysis used in accordance with the actual grade.

| No | Grade | Sample Quantity |
|----|-------|-----------------|
|    |   | TS | VS | Total |
| 1  | D1  | 22 | 10 | 32   |
| 2  | FANN| 22 | 10 | 32   |
| 3  | PFI | 22 | 10 | 32   |
The results of the NIR reflection spectra shown in Figure 1 are the results of 96 tea sample measurements. The spectral patterns are similar to one another because the tea varieties used have common similarities. However, the baseline and the intensity of each spectrum is varied. This is due to the influence of the used sample particles which have different textures that generate varying light scattering [13]. The scattering by different sizes of particles in the diffused-reflection mode of NIR spectral acquisition cannot be avoided. However, 1st derivative spectra are generally applied to bring all spectra into zero baselines [14].

Figure 1. Reflection spectra of 96 black tea samples.

Figure 1 shows the NIR reflectance spectra of 96 black tea samples. This reflectance spectrum contains not only signal from samples but also background and noise. To separate the signal from noise and background for PCA analysis, a pre-process treatment was applied including smoothing and derivations. Smoothing avoids local maximum and minimum artifacts and suppresses noise during spectral derivations, while 2nd derivative spectra improve peaks positions and spectral resolutions [15]. The smoothing was performed using Savitzky-Golay’s method before the 1st and 2nd derivative processes. The results of 2nd derivative spectra are shown in Figure 2.

Figure 2. The NIR 2nd derivative spectra of 96 black tea samples.
Figure 2 shows that the spectra are not completely overlapped both in positions and relative intensity. Close examinations showed several peak positions differed up to several tens of wavenumber, indicating slightly deference chemical dynamics are present from sample-to-sample. Several interactions between the chemical components present in the analysed material were well recognized such as the combination of the vibration of the CH molecular bonds in the energy span of 4545-4065 cm\(^{-1}\), the combination of NH and OH in the energy span of 5000 - 4545 cm\(^{-1}\), and the CH overtone in the energy span of 6173 - 5556 cm\(^{-1}\).

Tea grade classifications were performed using the NIR 2\(^{nd}\) derivative spectra as it successfully displayed different spectral characteristics among samples. PCA provides the principal component (PC) scores, which was the representations of the analyzed data matrix in PC space. PCA analysis on NIR 2\(^{nd}\) derivative spectra of 96 black tea samples showed the first three vectors of PC score in PC space, PC1, PC2, and PC3 successfully classified the training sets and validation sets of samples into three main clusters in which each cluster was practically occupied by the same grade of tea.

Figure 3. PCA plot for 3 grades of tea: D1 (red square), FANN (blue circle), PFI (blue triangle). The open symbol indicates the training set, and the filled symbol indicates the validation set. The percentage shown in each axis represents total variance explained.

Figure 3 shows three different clusters namely red squared (D1), blue circle (FANN), and the blue triangle (PFI). The open symbol is the training set, and the filled symbol is the validation set. Based on the PCA calculation for 66 training sets data, 60.01% of the variance can be explained by PC1, 21.76% by PC2 and 8.41% by PC3, resulted 90.18% variance can be explained using the above 3 PC’s. To validate the PCA classification method, 30 samples consisting of 10 samples for each grade were used as a blind sample test. The validation procedure was done by incorporated only one validation sample in every PCA analysis of validation to ensure that changes of the training sets parameters were minimum. The validation procedures were then repeated for all 22 validation samples. The overall results of the PCA analysis could differentiate 3 grades into PC1, PC2, and PC3, although a very few samples were misclassified. One D1 sample set falls in the FANN cluster, five FANN fall in the dust cluster, and one FANN data falls in the PFI cluster. Complete PCA analysis both for training sets and validation sets are summarized in Table 2.
Table 2. Results of the analysis using PCA on the 3 grades of tested samples.

| Grade | Number of Samples | No of Samples Classified by PCA |
|-------|-------------------|---------------------------------|
|       |                   | DUST | FANN | PF I |
| Training |                 |      |      |      |
| DUST   | 22                | 21   | 1    | 0    |
| FANN   | 22                | 5    | 16   | 1    |
| PF I   | 22                | 0    | 0    | 22   |
| Validation |             |      |      |      |
| DUST   | 10                | 10   |      |      |
| FANN   | 10                |      | 10   |      |
| PF I   | 10                |      |      | 10   |

Table 2 shows the percentages of the successful prediction by the PCA analysis reaching the accuracy of more than 91% (16 out of 22 samples were correctly classified) for the training sets and 100% (30 out of 30 samples were correctly classified) for the validation data. These results indicated that the method used in this study can be applied to classify the grade of black tea rapidly, accurately, and non-destructively.

4. Conclusion

A combination of Fourier-Transform Near-Infrared Diffusion-Reflection (FT-NIR) spectroscopy and a principal component analysis (PCA) can be used as an alternative method to examine the grade classifications of black tea with a rapid and non-destructive treatment. Tea samples of the same grade will clump together in the same classification, as in this study the samples were classified as PC1, PC2, and PC3. Although it is still necessary to add variations in training sets with different grades, this study showed that the general accuracy level of this classification method is over 90%.

Acknowledgments

The FSR was funded by an internal grant from Universitas Kristen Satya Wacana and Kemenristekdikti (grant 2016). The authors would like to thank Tafip Haryanto for preparing the NIR spectrometer prior measurements.

References

[1] International Tea Committee, 2015, World Tea Production & Consumption, The United Kingdom.
[2] Wibowo Z. S. and Sugiat, 1998, Indonesian tea between east and west, The 5th Int'l Symposium on tea culture pp 103-117.
[3] Abass T., Kustaniyati B. and Fa S., 2008, Rancang Bangun Pengendali Sistem Pengolahan Teh Hitam. 
[4] D Setyamidjaja, 2000 ,Teh: Budi Daya dan Pengolahan Pascapanen (Yogyakarta: Kanisius).
[5] Hartoyo A., 2003, Teh dan khasiatnya bagi kesehatan (Yogyakarta: Kanisius).
[6] Khorasani M., Amigo J. M., Sonnergaard J., Olsen P., Bertelsen P. and Rantanen J., 2015, Visualization and prediction of porosity in roller compacted ribbons with near-infrared chemical imaging (NIR-CI). J Pharm Biomed Anal. 109 11-17.
[7] Meng Y., Wan S., Cai R., Jiang B. and Zhao W., 2015, Discrimination and content analysis of fritillaria using near-infrared spectroscopy J. Anal. Methods Chem. 2015 1-8.
[8] Chang C. W., Laird D. A., Mausbach M. J. and Hurburgh C. R., 2001, Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. Soil Sci. Soc. Am. J 65(2) 480-90.
[9] Fahri N., Purwanto Y. A. and Budiasta I. W., 2017, Penggolongan Mangga Gedong Gincu
Berdasarkan Rasio Kandungan Gula Asam Menggunakan Prediksi Near Infrared Spectroscopy. Jurnal Keteknikan Pertanian 4(1) 31-36.

[10] Johnson R. A. and Wichern D. W., 2002, *Applied multivariate statistical analysis*, New Jersey: Prentice Hall 5(8).

[11] Yan S. -H., 2007, NIR evaluation of the quality tea and its market price, Spectroscopy Europe, 16(2) 16-19.

[12] Abdi H., 2010, Partial least squares regression and projection on latent structure regression (PLS Regression), *Wires Computational Statistics*, John Wiley & Sons Inc. 1-10.

[13] Tan S. M., Luo R. M., Zhou Y. P., Gong H. and Tan Z., 2012, Rapid and non-destructive discrimination of tea varieties by near-infrared diffuse reflection spectroscopy coupled with classification and regression trees *Afr. J. Biotechnol.* 11(9) 2303-12.

[14] Dryden G. M., 2003, *Near Infrared Reflectance Spectroscopy: Applications in Deer Nutrition* (Canberra: RIRDC Publication).

[15] Cen H. and He Y. 2007 Theory and application of Near Infrared Reflectance Spectroscopy in Determination of Food Quality *J. Trends in Food Sci. & Technol.* 18 72-83.