Applying LIBS-QCL Spectrum Coupled With Principal Component Analysis To Distinguish Gayo Arabica And Robusta Coffee

Rina Harvina Suci, Zulfahrizal, Agus Arip Munawar

Abstract— Laser Induced Breakdown Spectroscopy-Quantum Cascade Laser (LIBS-QCL) spectra acquisition for inorganic materials has been widely applied in various studies, but for the organic materials especially coffee has not been carried out. The objectives of this present study is to determine a spectral acquisition technique of LIBS-QCL for distinguishing Gayo Arabica and Gayo Robusta in the form of green beans, roasted beans, and coffee powder. Spectral data of coffee samples were acquired in wavelength range from 1000 to 2500 nm with co-added of 32 scans. Spectral data were enhanced by means of standard normal variation (SNV) and peak normalization (PN) algorithms. On the other hand, Principal Component Analysis (PCA) was applied in combination with LIBS spectrum as a method of data analysis and classification. The result showed that LIBS-QCL spectrum acquisition is quite good in the form of green beans. LIBS-QCL is able to distinguish the types of Arabica and Robusta coffee in the form of green beans and roasted beans, but coffee powder has not been able to be distinguished. The study also found several wavelength range intervals associated with coffee quality so that it can be used for further studies to develop coffee quality attributes prediction models.

Keywords— Non Destructive, Laser, Coffee, PCA, classification.

1 INTRODUCTION

Coffee is one of Indonesia's important export commodities as a source of foreign exchange besides oil and gas. One contributor to the export of Indonesia's coffee is Gayo Arabica coffee. Arabica coffee has a richer flavor character, that's why the Arabica coffee prices higher than Robusta coffee. Because the price of Arabica coffee in the market more expensive than Robusta coffee, it's feared it would be abused by persons who are not responsible [1]. One of them is by way of mixing between Arabica and Robusta aim to earn more profit. Mixing on foodstuffs particularly coffee can harm consumers and producers [2], [3]. Therefore, it needs a technology that can identify the mixing of Arabica and Robusta coffee varieties [4].

Rapidly and efficiently blending of Arabica and Robusta coffee mixing can be realized through the development of Laser-Induced Breakdown Spectroscopy (LIBS) technology using the Quantum Cascade Laser (QCL) spectra [5]. Detection of Arabica and Robusta coffee blending is done by classifying the coffee varieties based on the spectra of LIBS-QCL with Peak Normalization (PN) and Standard Normal Variate (SNV) pretreatment as a spectra correction method [6], [7]. LIBS-QCL considered to be suitable for determining the inner quality of foods and agricultural products since this method is characterized by low labour costs, simple sample preparation, non-destructive, pollution free and high speed of analysis [8], [9]. As a new method, LIBS-QCL research on organic materials has not been developed much, especially coffee plants. Therefore, it's hoped and expected that this technology can work rapidly and efficiently to distinguish Arabica and Robusta coffee in various forms.

2 MATERIALS AND METHODS

2.1 Sample Preparation

The sample used in this study was Gayo Robusta and Gayo Arabica coffee which had been given a post-harvest treatment. The total weight of the sample used was 350 grams Arabica coffee beans, 250 grams Robusta coffee beans, 210 gram Arabica roasted coffee beans, 150 grams Robusta roasted coffee beans, 122.87 grams Arabica coffee powder, and 117.15 grams Robusta coffee powder. In this study also used a blend of Arabica and Robusta varieties on each form of coffee with the blend composition were 75% of Arabica and 25% of Robusta.

2.2 Diffuse Reflectance Spectra Acquisition

Diffuse reflectance spectra in LIBS-QCL wavelength range of 1,000 - 2,500 nm were acquired for a total of 90 coffee samples. Work tool control for creating and running workflow using thermo integration software. Workflow created to set the tool to acquire diffuse reflectance spectra, and then the sample scanned 32 times, with frequency range in 573 ± 18 Hz, the optical attenuation 4 times, and a resolution of 8 cm⁻¹, then the results were averaged and saved in three different file formats (*.SPA, *.JDX and *.CSV) Furthermore, the spectra data is processed using unscremlar software ® X version 10.1 Nntwork client [10], [11].

2.3 Outlier Removal

Spectra data were firstly projected onto Principal Component Analysis (PCA) followed by Hotelling $T^2$ ellipse for outlier detection. If there were any data outside the ellipse, then these spectra data were noted as outlier and removed prior to further analysis [12], [13].

2.4 Coffee classification

Corrected and enhanced LIBS-QCL spectra data were subjected to PCA in order to identify and classify samples based on their similarities. PCA employs a mathematical procedure that transforms a set of possibly correlated response variables into a new set of non-correlated variables [14], [15].
2.5 Spectra data correction

In order to obtain accurate and robust classification results, spectra data were enhanced and corrected using these following methods: Peak Normalization (PN), Standard Normal Variate (SNV). Spectra correction tends to remove spectral noises due to light scattering, sensor curvature, over-heated effects, and multiplicative effects [16], [17].

3 Results and Discussions

3.1 Typical spectra of coffee samples

A Typical diffuse reflectance spectrum for coffee beans, roasted coffee beans, and coffee powder was presented in Fig.1. It shows several peaks represent the vibration of molecular bonds of C-C, O-H, N-H, C-H-O and C-H. Original spectra data before correction were still interference due to noise resulted from light scattering. These noises were corrected using several pre-treatment methods such as peak normalization and standard normal variate. Spectra correction methods clearly enhance spectra appearance and remove some noises due to light scattering.

![Diffuse reflectance spectra of coffee beans, roasted coffee beans, and coffee powder sample](image)

Fig. 1. Diffuse reflectance spectra of coffee beans, roasted coffee beans, and coffee powder sample

When the light goes through coffee samples, there are several reactions given by the samples. Some of light are reflected, absorbed and transmitted. The different reactions depend on main chemical composition, cell structure and other chemical properties of the samples. As shown in Fig. 1, amino acid and chlorogenic acid content of coffee beans, roasted coffee beans, and coffee powder samples are observed in wavelength area between 2010-2071 nm and 2040-2225 nm; whilst absorption bands at around 1911-1938 nm is associated with water content; lipid and caffeine content are observed in wavelength area between 1730-1763 nm and 1677-1718 nm. We may argue that amino acid, chlorogenic acid, water content, lipid and caffeine concentrations can be predicted in those wavelength region [4].

The spectra characteristic of these three coffees has a similar shape, it shows though different forms, but it has a relatively similar chemical composition. LIBS-QCL spectra of coffee is essentially composed of a large set of overtones and combination bands and further may be complicated since the spectra is influenced by wavelength dependent scattering effects, tissue heterogeneities, instrumental noise, ambient effects and other source of variability [18]. These factors may generate spectra noise and influence LIBS-QCL classification performance. Several methods are introduced as spectra pre-treatment to overcome these factor effects. Thus, spectra pre-treatment was performed prior to further chemometrics analysis. Peak Normalization (PN) and standard normal variate (SNV) were used for this purposes [5].

3.2 Outlier Removal

Firstly, spectra data of coffee beans, roasted coffee beans, and coffee powder samples were analyzed using principal component analysis to inspect any dissimilarities. Hotelling T^2 ellipse was projected onto PCA map to detect outlier data. Data outside the ellipse were noted and removed due to the potential of outliers to affect the accuracy and robustness of the classification model [6]. Outlier data analysis was present in Fig. 2.

![Outlier Data Analysis on Coffee Beans](image)

Fig. 2a. Outlier Data Analysis on Coffee Beans

![Outlier Data Analysis on Roasted Coffee Beans](image)

Fig. 2b. Outlier Data Analysis on Roasted Coffee Beans

As can be seen in Fig.2a, there was 1 lever data found since the data were outside the ellipse. Thus, all the total of 10 spectra data were used to classify coffee beans. The same thing as coffee beans, there was 1 lever data found in roasted coffee beans as shown in Fig.2b whilst 1 outlier data was found in coffee powder. Therefore, 9 spectra data were used to classify coffee powder as shown in Fig.2c.

3.3 Coffee Beans Classification Using Principal Component Analysis (PCA)

Corrected and enhanced LIBS-QCL spectra data were subjected to PCA in order to identify and classify samples based on their similarities. PCA employs a mathematical procedure that transforms a set of possibly correlated response variables into a new set of non-correlated variables. Classification with PCA method were presented in Fig. 3.
In this study, classification was successfully accomplished by using two principal components (PCs) with a total explained variance 99% for each varieties (Arabica and Robusta). As shown in Fig.3, the data is well classified with a success percentage of 90% for Arabica varieties and 75% for Robusta varieties.

3.4 Spectra Data Correction for Coffee Beans

The classification results of Arabica and Robusta varieties in coffee beans need to be repaired to obtain a better classification. To enhance and improve classification performance, spectra correction was employed to all spectra data of coffee samples. Peak Normalization (PN) and Standard Normal Variate (SNV) was chosen as spectra data correction method as shown in Fig. 4.

As can be seen in Fig.4, PCA Classification with the addition of PN pre-treatment resulted in total (PC-1 and PC-2) 99% for each Arabica and Robusta varieties. This data is well classified by 90%.

As can be seen in Fig.5 above, The PCA classification with the addition of SNV pretreatment resulted in a total explained variance 89% for each Arabica and Robusta varieties. This data is well classified by 90%.

3.5 Roasted Coffee Beans Classification Using Principal Component Analysis (PCA)

PCA classification without pre-treatment was successfully accomplished by using two principal components (PCs) with a total explained variance 100% for each varieties (Arabica and Robusta). Classification with PCA method on roasted coffee beans were presented in Fig. 6.
3.6 Spectra Data Correction for Roasted Coffee Beans

PCA Classification with the addition of PN pre-treatment resulted in total (PC-1 and PC-2) 100% for each Arabica and Robusta varieties. The classification with the addition of PN pre-treatment can be seen in Fig. 7.

Based on Fig. 7, the data is well classified by 80% for Arabica and 75% for Robusta.

Based on Fig. 8 above, PCA classification with the addition of SNV pretreatment resulted in a total explained variance 94% for each Arabica and Robusta varieties. This data is well classified by 90% for Arabica, and 75% for Robusta.

3.7 Coffee Powder Classification Using Principal Component Analysis (PCA)

PCA analysis without pre-treatment was successfully accomplished by using two principal components (PCs) with a total explained variance 100% for each variety (Arabica and Robusta). Classification with PCA method on roasted coffee beans were presented in Fig. 9.

This data is well classified with a success percentage of 78.85% for Arabica varieties and 80% for Robusta varieties.
3.8 Spectra Data Correction for Coffee Powder

PCA Classification with the addition of PN pre-treatment resulted in total (PC-1 and PC-2) 99% for each Arabica and Robusta varieties. The classification with the addition of PN pre-treatment can be seen in Fig. 10.

![Fig. 10a. PCA classification of Arabica varieties with PN pre-treatment](image)

![Fig. 10b. PCA classification of Robusta varieties with PN pre-treatment](image)

Based on Fig. 10 above, the data is well classified by 80% for Arabica and 75% for Robusta.

![Fig. 11a. PCA classification of Arabica varieties with SNV pre-treatment](image)

![Fig. 11b. PCA classification of Robusta varieties with SNV pre-treatment](image)

Based on Fig. 11 above, the PCA classification with the addition of SNV pretreatment resulted in a total explained variance 94% for each Arabica and Robusta varieties. This data is well classified by 90% for Arabica, and 75% for Robusta.

PCA results proved that LIBS-QCL was able to classify data based on Arabica and Robusta varieties. But from the classification result using PCA method in the form of coffee powder with the addition of pre-treatment or without pre-treatment has not been able to classify the data. PCA classification that given pre-treatment was also no better than PCA classification without pre-treatment. We can conclude that coffee powder was not successfully classified.

This can be caused by the magnitude of the laser energy emitted to the material that is suspected to damage the chemical bonds within the material itself. Unlike the coffee beans and roasted coffee beans, coffee powder no longer has a hard skin on the seeds that can protect it from the amount of laser energy radiated. This statement was also in agreement with Hooker Simon (2010) from what they said that Laser (Light Amplification by Stimulated Emission of Radiation) is a source that can emit light with almost perfect levels of monochromatism. The laser beam differs from the ordinary light. There are some characteristics of a laser that distinguishes with ordinary light, which is merely radiating on one direction only (linear polarized), has a high intensity, almost perfect monochromatism and has a high level of coherence.

In previous studies, LIBS-QCL has never been tested to organic materials. So far only applied to inorganic materials. Therefore, this study tries to test on organic matter. Although not all study objectives are achieved, but some objectives are achieved. Therefore, further research is needed in order to ensure that LIBS-QCL can be used in organic materials.

In general, it can be concluded that the PCA method is able to classify coffee based on its varieties where all PCA classification models use pretreatment and without pretreatment obtain models with a success rate above 75% except in the form of coffee powder that has not been able to be classified. Based on table 1, it can be concluded that the best pre-treatment used to distinguish Arabica coffee and Robusta coffee is SNV pre-treatment with 86.58% success. Coffee classification results in various forms was shown in Table 1.
Table 1. Coffee classification results in various forms.

<table>
<thead>
<tr>
<th>Pretreatment</th>
<th>Coffee Beans</th>
<th>Roasted Coffee Beans</th>
<th>Coffee Powder</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>R</td>
<td>A</td>
<td>R</td>
</tr>
<tr>
<td>Raw</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>75</td>
</tr>
<tr>
<td>PN</td>
<td>90</td>
<td>90</td>
<td>80</td>
<td>75</td>
</tr>
<tr>
<td>SNV</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>75</td>
</tr>
</tbody>
</table>


4 CONCLUSION

Based on obtained results, we may conclude that LIBS-QCL Spectrum Acquisition Technique for the best results is obtained by coffee beans at a frequency range of 573 ± 18 Hz, scan number 32 times, optical attenuation as much as 4 times, resolution of 8 cm⁻¹, and at wavelength region 1000 - 2500 nm. LIBS-QCL technology combined with PCA method is able to identify and classify coffee beans and roasted coffee beans based on their varieties (Arabica and Robusta) while coffee powder is not yet able to be distinguished. Standard normal variate (SNV) found to be the best spectra enhancement method in classifying Arabica and Robusta coffee.

REFERENCES


