PREDICTION OF CAFFEINE CONTENT IN LIBERICA COFFEE GREEN BEAN BY NIR SPECTROSCOPY USING KUBELKA-MUNK MODEL

PENDUGAAN KANDUNGAN KAFEIN BIJI KOPI LIBERIKA DENGAN SPEKTROSKOPI NIR MENGGUNAKAN MODEL KUBELKA-MUNK

* Ei Mon Kyaw1), I W Budiastra2), Sutrisno3), Samsudin3), Dheni Mita Mala4)
1) Postharvest Technology Study Program, Postgraduate, IPB University
   Jl. Raya Dramaga Campus IPB Dramaga, Bogor 16680, Indonesia
2) Department of Mechanical and Biosystem Engineering, IPB University
   Jl. Raya Dramaga Campus IPB Dramaga, Bogor 16680, Indonesia
3) Indonesian Industrial and Beverage Crops Research Institute (BALITRI)
   Jl. Raya Pakuwon Km 2 Parungkuda, Sukabumi 43357 Indonesia
4) Center of Agro-based Industry (BBIA), Ministry of Industry Republic of Indonesia
   Jl. Ir. H. Juanda No.11, RT.04 / RW.02, Paledang, Bogor 16122, Indonesia
   *eimonlay.horti@gmail.com

(Tanggal diterima: 2 Juni 2020, direvisi: 16 Juli 2020, disetujui terbit: 1 September 2020)

ABSTRACT

Liberica is one of coffee species that is becoming popular and increasingly in demand in present days due to its unique characteristics. Caffeine is one of the important coffee quality parameter which determines the coffee flavor, consumer preference and market price. Caffeine content is usually analyzed by chemical method which is destructive, time consuming, expensive and involving a lot of procedures. NIR Spectroscopy is one of the non-destructive techniques to overcome these disadvantages. This study was conducted at the Department of Mechanical and Biosystem Engineering, IPB University for NIR measurement and the Center of Agro-based Industry (BBIA), Bogor for chemical analysis from August to November 2019. The study aimed to determine the best calibration model for the prediction of caffeine content in Liberica coffee green bean powder. In this study, FT-NIRS in the wavelength of 1000-2500 nm was used for NIR measurement and HPLC tool was used for chemical analysis. Kubelka-Munk (K/S) and Absorbance (Log 1/R) were used as data transformation, whereas Standard Normal Variance (SNV) and Second derivative of Savitzky-Golay (dg2) as data pretreatment. In addition, Partial Least Square (PLS) and Multiple Linear Regression (MLR) were applied for multivariate calibration method. The best calibration model for the prediction of caffeine content of Liberica coffee green bean powder was obtained by the spectral data pretreated with second derivative of Savitzky-Golay (dg2) and Kubelka-Munk data transformation using PLS calibration method with the results of $r = 0.90$, RPD = 2.24, CV = 2.01%.

Keywords: Caffeine content; coffee; Kubelka-Munk; Liberica; multivariate calibration
INTRODUCTION

Coffee is one of the most traded agricultural products worldwide. There are a lot of coffee species grown throughout the world. However, Arabica (Coffea arabica L.) and Robusta (Coffea canephora L.) are most commonly found in global trade markets. Meanwhile, Liberica coffee (Coffea liberica) species began to gain popularity in recent years. One unique characteristic of Liberica is its adaptability to peatlands and its tolerance to disease attacks. Liberica coffee (Coffea liberica) is found mostly in Malaysia, Guyana, Liberia, Suriname, Philippines, Equatorial Guinea, Nigeria, Sao Tomé and Comoros (Min, 2016; Old Courthouse, 2019). In Indonesia, Liberica coffee plantations can be found in Sumatra, especially in Tanjung Jabung Barat, Jambi Province (Kirana & Utami, 2018). Sianipar (2017) stated that currently Liberica coffee is better known by consumers and the demand for Liberica coffee beans tends to increase in the present. This is evidenced by the high price of Liberica coffee exports from Indonesia to Malaysia (Rp.33,000 – 40,000/kg at the farm level).

Liberica is also known as jackfruit coffee due to its flavor and aroma. Coffee quality is mainly determined by the flavor and aroma which is influenced by the chemical compounds in the coffee bean, such as caffeine. Caffeine has bitter flavor and physiologically stimulating in human. Buffo & Cardelli-freire (2018) stated that caffeine contributes to the strength, body and bitterness of brewed coffee. Caffeine contents in green bean are different in each coffee species; 1.1-1.3% caffeine in Arabica, 2.4-2.5% in Robusta and 1.1 to 1.3% in Liberica (Farah, 2012; Hulupi & Martini, 2013). Shiferaw & Alemanyu (2018) mentioned that high caffeine content adversely affects the price of coffee and consumer preferences to coffee products (beans). Therefore it is necessary to determine caffeine contents of coffee beans before they are released into the beans market or exported.

Determining chemical contents is usually carried out by using chemical method which is expensive, time consuming and long procedures. Near Infra-Red (NIR) Spectroscopy is one of the non-destructive methods and widely used for quality assessment of agricultural products due to its rapidity, simplicity, cost effectiveness, and safety. Purningsih (2018) stated that the Kubelka-Munk method showed the relationship between reflectance and absorbance and the scatter of light that occurs. Kubelka-Munk is used for samples with small particles that connect the spectra reflectance of the sample, absorption (K), and scattering characteristics (S) so that the use of the Kubelka-Munk equation is considered more effective to increase the linearity of the spectrum of chemical data especially in samples with small particles (powder).

Numerous NIRS studies predicting chemical contents of coffee products, yet NIRS prediction for Liberica coffee green bean has not been carried out. In this study, the Standard Normal Variance (SNV) and second derivative of Savitzky-Golay (dg2) as data pretreatment was implemented, as well as Kubelka-Munk (K/S) method and Absorbance (Log 1/R) as data transformation. Partial Least Square (PLS) method and Multiple Linear Regression (MLR) were used for calibration. This research aimed to predict the caffeine content of Liberica coffee green bean powder by using NIR Spectroscopy and to determine the best calibration model, data transformation and data pretreatment for caffeine content of Liberica coffee.
MATERIALS AND METHODS

The study was conducted at the Department of Mechanical and Biosystem Engineering, IPB University for NIR measurement and the Center of Agro-based Industry (BBIA), Bogor for chemical analysis, from August to November 2019.

Materials and Sample Preparation

The materials used in this study were Liberica coffee green beans from Pakuwon Experimental Station of Indonesian Industrial and Beverage Crops Research Institute (BALITTRI) in Sukabumi, which is located at an altitude of 450 meters above sea level, and coordinates at 6° 49’58.0" S and 106° 44’28.4" E.

Liberica coffee green beans (9.9% water content) were ground by a disk mill and filtered by using sieve (mesh No.100). Fifty samples of coffee powder were used for NIR spectroscopy measurement. Two third of samples (34 samples) were taken for calibration model development and 1/3 of samples (16 samples) for validation model.

Sample Measurement

Fifty grams of coffee powder were spread into the petri dish until covering 1.2 cm of the petri dish height. NIR measurement was then conducted to these samples by Spectrometer type NIRFlex N-500 with the wavelength of 1000 – 2500 nm and the scan speed of 3 scans per second (temperature at 22 – 25°C). A total of 150 spectral data were obtained.

After NIRS measurement, the caffeine content analysis was executed using HPLC tools (Shimadzu LC-20AD) (Rai et al., 2015). Ten grams of coffee green bean powder for each sample was added to 100 ml of hot purified water at 80°C. After stirring for one minute, the mixture was centrifuged at approximately 240 rpm. The supernatant was filtered with 0.45 μm filter and was applied to HPLC with minor corrections, using an Inert Sustain C18 column at 40°C. A mobile phase consisted of 0.1% phosphoric acid and 20% acetonitrile. Elution was carried out at a flow rate of 1.0 ml/min and the results were detected at 320 nm. The concentration of caffeine in the sample \(C_{caf}\) was calculated using the regression equation and compared with standard solutions. The percentage of caffeine was calculated by Equation (1):

\[
\text{Caffeine (\%)} = \frac{C_{caf} \times V_{sol}}{W_{sample}} \quad (1)
\]

Where:

- \(C_{caf}\) = concentration of caffeine in the sample (%)
- \(V_{sol}\) = volume of solution (ml)
- \(W_{sample}\) = weight of sample (g)

NIR Data Processing and Analysis

The reflectance spectrums were transformed into absorbance (log 1/R) and Kubelka Munk Model (K/S). In this study, standard normal variance (SNV) and second derivative (dg2) were applied as NIR data pretreatment. The samples were used for calibration (102 samples) and validation (48 samples). The calibration and validation for NIR data processing and chemical contents were carried out using Partial Least Square (PLS) and Multiple Linear Regression (MLR). Accuracy of NIRS method for caffeine content prediction of Liberica coffee was determined by the result of Coefficient of Correlation (r), Standard Error of Calibration (SEC), Standard Error of Validation (SEV), Coefficient of Variation (CV), Ratio of Performance to Deviation (RPD) and its consistency. NIR data processing and analysis were carried out using Unscramble software V10.5 trial version (CAMO, Norway).

RESULTS AND DISCUSSION

Original Reflectance Spectrum

The original reflectance spectrum of Liberica coffee green beans powder at the wavelengths of 1000-2500 nm can be seen in Figure 1. There are 1501 spectral data in the wavelength range of 1000-2500 nm and has a lot of peak and valley of the spectra. The existence of peaks and valleys of the spectrum is influenced by the chemical components which are contained in the material being analyzed. Purningsih (2018) mentioned that the spectral peaks of caffeine can be observed at the wavelength 1128 nm, 1298 nm, 1672 nm, 1726 nm and 1934 nm. Moreover, Rosita (2016) stated that there are four main spectral peaks of caffeine at the wavelength of 1128 nm, 1298 nm, 1672 nm and 1726 nm in the NIR spectrum of Arabica Gayo coffee green bean.
Figure 1 showed the spectral peak and valley for caffeine content of Liberica coffee green bean powder at 1298 nm and 1672 nm as spectral peak, whereas 1726 nm and 1934 nm as valleys. However, the wavelength of 1128 nm did not appear in this study due to the original spectral data attained from NIR instrument cannot show all information of the material. It also has some interference effect on the relevant information such as noise, background and hidden information due to overlapping information with another (Purningsih, 2018). It is necessary to eliminate or reduce noise, hidden and background information in order to obtain reliable, accurate and stable calibration models. Therefore, spectral data pretreatment prior to modeling is vital.

Caffeine Content of Liberica Coffee Green Bean Powder

The caffeine content of Liberica coffee green bean powder is shown in Table 1. The results of chemical analysis were used as reference data for calibration and validation. Rosita, Budiastra & Sutrisno (2016) mentioned that the highest and the lowest values in the processed NIRS data are intended as calibration data sets in building the calibration model.

The caffeine content of Liberica coffee green beans powder found in this study ranged between 1.12% and 1.35% with the standard deviation of 0.055% for calibration and 0.021% for validation (Tabel 1). Meanwhile previous study conducted by Purningsih (2018) showed that caffeine content in Bondowoso Arabica coffee green bean powder was 1.27% - 1.57%, Arabica Java Preanger coffee bean was 1.18 – 1.41 % (Budiastra, Sutrisno, Widyotomo, & Ayu, 2018) and Arabica Gayo coffee beans was 0.194 to 0.328% (Rosita, Budiastra & Sutrisno, 2016). Hulupi & Martini (2013) stated that caffeine content of Liberica coffee green bean was 1.1% to 1.3%. Therefore, there is no difference in the caffeine content of Liberica coffee green bean and Liberica coffee green bean powder.

Calibration and Validation of Liberica Green Bean Powder

The results of calibration and validation model using PLS method are shown in Table 2. The best calibration model can be found in the spectra, which pretreated with second derivative of Savitzky-Golay (dg2) and by Kubelka-Munk data transformation, shown by the correlation coefficient r (0.90), RPD (2.24), consistency (80%) and CV (2.01%). Savitzky-Golay (dg2) can eliminate background information and increase the resolution of the spectra. Furthermore, Kubelka-Munk data transformation method affects the
relationship between reflectance and absorbance as well as the scatter of light.

All spectral data except dg2 transformed by Absorbance (Log 1/R) in Table 2 also showed good accuracy for the caffeine content prediction in Liberica coffee green bean powder. However, spectral data pretreated with dg2 transformed by Absorbance (Log 1/R) showed poor accuracy because Absorbance (Log 1/R) data transformation method did not affect light scattering of small particles samples (coffee powder). Table 2 exhibited that the best calibration model can be found mostly in the spectral data with Kubelka-Munk data transformation because the method is based on absorption and reduced scattering effect in the spectra of small particles, such as in powder form. This result was in line with the study conducted by Purningsih (2018), which stated that spectral data transformed by Kubelka-Munk method combined with dg2 showed better accuracy than absorbance data (Log 1/R) in predicting caffeine content in green bean powder.

Table 3 showed that the results of calibration and validation using MLR method. MLR is a combination of the wavelength of the main chemical content and several wavelengths that accommodate the effect of scatter and intercorrelation between chemical components. Furthermore, MLR method determines the wavelength that is dominant in predicting chemical content. According to Athfiyah (2017), wavelength selection method can be used to improve model performance without spectra data processing. Therefore, the calibration process was done using reflectance spectral data and the original spectral data transformed by Kubelka-Munk method. The best result was obtained through the use of 10 selected wavelengths in Kubelka-Munk spectral data with r-value of 0.82, consistency of 108%, RPD of 2.10 and CV of 2.14%. This result is better than the previous study done by Budiastra, Sutrisno, Widyotomo, & Ayu (2018), which used 10 selected wavelength with r-value of 0.70, consistency of 90%, RPD of 1.21 and CV of 3.23% in predicting caffeine content in Java Preanger coffee bean. The use of coffee powder in predicting caffeine content showed good accuracy by MLR method. The remaining model could not be used to predict caffeine content accurately. The best mathematical model for caffeine content with the selected wavelength is shown in Equations 2.

### Table 2. Result of calibration and validation for caffeine content by PLS (Partial Least Square)

| Data Transformation | Data Pretreatment | Factor | r   | SEC (%) | SEV (%) | CV (%) | RPD  | Consistency (%) |
|---------------------|-------------------|--------|-----|---------|---------|--------|------|-----------------|
| Absorbance          | Original          | 10     | 0.89| 0.0224  | 0.0276  | 1.84   | 2.15 | 81              |
|                     | SNV               | 8      | 0.83| 0.0266  | 0.0280  | 2.19   | 1.80 | 95              |
|                     | Dg2               | 1/p6   | 0.65| 0.3823  | 0.4291  | 3.37   | 1.32 | 87              |
| Kubelka Munk        | Original          | 10     | 0.85| 0.0258  | 0.0314  | 2.13   | 1.92 | 82              |
|                     | SNV               | 10     | 0.84| 0.0294  | 0.0328  | 2.42   | 1.86 | 90              |
|                     | Dg2               | 4/p4   | 0.90| 0.0434  | 0.0305  | 2.01   | 2.24 | 80              |

Notes: r = coefficient of correlation; SEC = standard error of calibration; SEV = standard error of validation; CV = coefficient of variation; RPD = ratio of performance to deviation; SNV = standard normal variance; dg2 = second derivative of Savitzky-Golay

Keterangan: r = koefisien korelas; SEC = galat baku kalibrasi; SEV = galat baku validasi; CV = koefisien keragaman; RPD = rasio kinerja terhadap simpangan; SNV = ragam normal baku; dg2 = turunan kedua dari Savitzky-Golay

### Table 3. Result of calibration and validation for caffeine content by MLR (Multiple Linear Regression)

| Data Transformation | Wavelength number | r   | SEC (%) | SEV (%) | CV (%) | RPD  | Consistency (%) |
|---------------------|-------------------|-----|---------|---------|--------|------|-----------------|
| Reflectance         | 10 WL             | 0.74| 0.0335  | 0.0418  | 2.76   | 1.48 | 80              |
|                     | 5 WL              | 0.63| 0.0388  | 0.0454  | 3.20   | 1.28 | 85              |
| Kubelka Munk        | 10 WL             | 0.82| 0.0260  | 0.0240  | 2.14   | 2.10 | 108             |
|                     | 5 WL              | 0.79| 0.0280  | 0.0212  | 2.31   | 2.01 | 132             |

Notes: r = coefficient of correlation; SEC = standard error of calibration; SEV = standard error of validation; CV = coefficient of variation; RPD = ratio of performance to deviation

Keterangan: r = koefisien korelas; SEC = galat baku kalibrasi; SEV = galat baku validasi; CV = koefisien keragaman; RPD = rasio kinerja terhadap simpangan
Figure 2. Plot of caffeine content referenced vs. predicted using PLS (Partial Least Square) method

Gambar 2. Plot kandungan kafein referensi vs. prediksi menggunakan metode PLS (Kuadrat Terkecil Parsial)

\[
Y = 0.38 + 74.19 K/S_{2951.58} + 56.67 K/S_{2355.15} + 137.67 K/S_{2263.14} - 161.01 K/S_{2079.05} + 208.13 K/S_{1956.04} - 551.37 K/S_{1847.75} + 274.66 K/S_{1842.3} - 617.71 K/S_{1786.99} + 202.85 K/S_{1784.44} + 897.636 K/S_{1663.34}
\]

Where:
- \( Y \) = Caffeine content
- \( K/S_n \) = Kubelka-Munk from wavelength number ‘n’

Figure 2 and 3 showed the scatter plot of reference and prediction values of caffeine content by using PLS and MLR method, respectively. These two figures showed a correlation and RPD between caffeine content predicted using NIRS and caffeine reference using HPLC, indicated by \( r \)-value > 0.80 and RPD > 1.5. This result showed that the obtained calibration models using both PLS and MLR method could predict caffeine content in Liberica coffee green bean. However, the calibration model using PLS method showed better accuracy compared to MLR method because MLR method used only the spectral data without any data pretreatment, which reduces baseline shift and interference of the chemical components in coffee.
CONCLUSIONS

This study proved that NIR Spectroscopy determines the caffeine content of Liberica coffee. The best calibration model for the prediction of the caffeine content of Liberica coffee green bean powder was obtained by the spectral data pretreated with second derivative of Savitzky-Golay (dg2) and Kubelka-Munk data transformation using PLS calibration method. The result showed r-value of 0.90, RPD of 2.24, and CV of 2.01%. Kubelka-Munk (K/S) method is more reliable in data transformation of small particles as the coffee powder has a scattering effect. PLS is more recommended in predicting caffeine content in Liberica coffee green beans.

ACKNOWLEDGMENT

The authors would like to thank the Indonesian Industrial and Beverage Crops Research Institute, Sukabumi for the coffee processing, Center of Agro-based Industry (BBIA) for caffeine content analysis and TPPHP laboratory, Mechanical and Biosystem Engineering Department, IPB University.

AUTHORS CONTRIBUTION

1. Ei Mon Kyaw (Main contributor)
2. I.W. Budiastra (Member contributor)
3. Sutrisno (Member contributor)
4. Samsudin (Member contributor)
5. Dheni Mita Mala (Member contributor)

REFERENCES

Athfiyah, S.N. (2017). Penentuan kandungan kimia utama biji kopi Arabika Bondowoso secara non destruktif dengan Near Infrared Spectroscopy. Institut Pertanian Bogor, Bogor.

Budiastra, I. W., Sutrisno, Widiotomo, S., & Ayu, P. (2018). Prediction of Caffeine Content in Java Preanger Coffee Beans by NIR Spectroscopy Using PLS and MLR Method, 147 (2018), https://doi.org/doi:10.1088/1755-1315/147/1/012004 1234567890

Buffo, R. A., & Cardelli-freire, C. (2018). Coffee flavour: An overview Coffee Flavour: an Overview, (February). https://doi.org/10.1002/ffl.1325

Farah, A. (2012). Coffee: Emerging Health Effects and Disease Prevention (pp. 21–38). Blackwell Publishing Ltd.

Hulupi, R., & Martini, E. (2013). Budi Daya dan Pemeliharaan Tanaman Kopi di Kebun Campur. Bogor, Indonesia.

Kirana, C. A., & Utami, D. T. (2018). The Effect of Liberica Coffee on Liver Physiology and Histological Appearance of White Mice, (October), 152–159.

Min, T. H. (2016). Phytochemicals and Antioxidative Properties of Coffea Liberica Green Beans and Comparison with that of Coffea Arabica and Coffea Robusta, (December)

Old Courthouse, (2019). Liberica Coffee International Roasting Competition 2019, (April), Malaysia

Purningsih, V., Budiastra, I. W., Sutrisno, & Widiotomo, S. (2018). Akurasi metode NIRS dalam prediksi kandungan kimia green kopi Arabika Bondowoso dengan model Kubelka-Munk. DOI: 10.19028/jtep.06.3.271-278, Vol. 6 No. 3, p 271-278

Rai, K. A., Erashima, H. T., Izawa, S. A., Aga, A. T., Amamoto, A. Y., Sutsumiuchi, K. T., & Odama, S. K. (2015). Simultaneous Determination of Trigonelline , Caffeine , Chlorogenic Acid and Their Related Compounds in Instant Coffee Samples by HPLC Using an Acidic Mobile Phase Containing Octanesulfonate, JI(August), 831–835.

Rosita, R. (2016). Penentuan kandungan kimia biji kopi arabika Gayo secara non destruktif dengan Near Infrared Spectroscopy. Institut Pertanian Bogor, Bogor.

Rosita R, Budiastra IW, Sutrisno. 2016. Prediction of Caffein Content of Arabica Coffee Bean by Near Infrared Spectroscopy. JTEP Jurnal Keteknikan Pertanian. Vol. 4 No. 2, p 179-186

Shiferaw, M., & Alemayehu, I. (2018). Determination of Caffeine Content of Bale Coffee Using HPLC Analysis, 73, 23–32.

Sianipar, H. (2017). Keragaman genetik populasi kopi Liberika (coffea liberica w. Bull ex. Hiern) di Kecamatan Betara berdasarkan karakter buah dan biji. Universitas Jambi, Jambi.
