NIR reflectance spectroscopy and SIMCA for classification of crops flour

R E Masithoh1 and I Yuliyanda1

1 Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Indonesia

E-mail: evi.ugm@gmail.com

Abstract. The potential of SIMCA technique for crops flour classification was studied based on FT-NIR spectroscopy in this research. A total of 72 spectra of flour samples taken from 6 types of crops, i.e. of banana, breadfruit, taro, arrowroot, purple sweet potato, and modified cassava (mocaf). The reflectance data were measured using the NIRflex N500 Fiber Optic Solids Cell at 4000 – 10,000 cm⁻¹. The spectral obtained were pre-processed and analyzed using The Unscrambler X version 10.5.1. A 2nd derivative Savitzky-Golay (polynomial order 2, 25 smoothing points) followed by a Standard Normal Variate (SNV) were used for pre-treatment data. Characterization of the flours was done using chemometric models based on principal component analysis (PCA) and soft independent modeling of class analogy (SIMCA) explained for each group of flour samples of banana, breadfruit, taro, arrowroot, purple sweet potato, and modified cassava (mocaf). SIMCA calibration models were constructed using 6 spectral measurements for each type of flours; classification set were constructed using 6 spectral measurements. The SIMCA accuracy classification were 100% for mocaf, banana, arrowroot, bread fruit, and taro, and 67% for purple sweet potato.

1. Introduction

Rice has been long known as staple food in Indonesia making rice as the primary source for carbohydrate. Recently, people is starting to put wheat based products, such as bread and noodle, in their diet to alternate rice. Rice and wheat are not only rich in carbohydrate but also high in glycemic index (GI) of 89 and 100, respectively [1]. GI is an index of how carbohydrate-rich foods raise the blood sugar levels, therefore the information of the GI is important in diet management and diabetes mellitus prevention. Among carbohydrate source, rice and bread from wheat are classified as high GI of 89 and 100, while GI of cassava, sweet potato, and taro are 74, 65, and 69 [1,2].

Tubers such as arrowroot (Marantha arundinacea), purple sweet potato (Ipomoea batatas), taro (Colocasia esculenta) and cassava are easily found but their use as food products are limited. Banana and breadfruit (Artocarpus altilis) are also abundantly available but most of them are consumed fresh or fried. Those crops are the main sources of starch that is a macronutrient in several foods and supplies majority of the calories consumed by most people in the world [3]. Apart from the starch contents, those crops have lower protein compared to wheat or rice [4-6]. It is a challenge to find energy source crops which have high carbohydrate with low GI but high protein.

Several analytical techniques are used in protein analysis, among which NIR spectroscopy for protein determination of wheat [7], rice [8,9], or grains [10]. NIR spectroscopy based on the absorption of electromagnetic energy between 400 and 14,000 cm⁻¹ provides a rapid and non-destructive analytical
technique. NIR techniques have been successful in identification or classification of fruits [11], beverages [12,13], adulterated products [14,15]. Unfortunately, only few studies conducted for local, low protein constituent crops.

Considering the lack of study of local crops, this study aimed at combining NIR spectroscopy and pattern recognition for various flours made of local crops. Principal component analysis (PCA) was used for unsupervised pattern recognition and self modeling independent cluster analysis (SIMCA) was used for supervised pattern recognition. This study employed several PCA models which result in the best model for flour classification.

2. Materials and Methods

2.1. Materials and Methods

2.1.1. Materials. Total sample for this research were 72 flour samples made of 6 types of crops such as banana, breadfruit, taro, arrowroot, purple sweet potato, and modified cassava flour purchased from local farmers. Each crops contributed 12 samples to accommodate sample variance.

2.1.2. Determination of protein. Protein content was determined using Kjeldahl method [15] which in short consisted of conversion wet digestion nitrogen to ammonium sulfate, neutralization to free ammonia, distillation ammonia into boric acid, titration with alkali, conversion nitrogen concentration to protein using conversion ratio, then finally determination of ‘true protein’ by precipitating out protein, analyzing remaining nitrogen, subtracting from total nitrogen content. All experiments were conducted in triplicate.

2.1.3. Spectral acquisition. Approximately 4 g of a sample was weighed and transferred to a glass sample container (with a height of 10 mm and a diameter of 2 cm). To maintain uniformity of sample during spectra measurement, the sample was pressed with a glass slide beforehand. Near-infrared reflectance spectra of the samples were obtained in the reflectance mode on a NIRFlex N500 Fiber Optic Solids Cell spectrometer (Büchi Labortechnik AG, Flawil, Switzerland) with InGaAs detector. Spectra were acquired in reflectance mode in the long-wavelength region of 4000 – 10,000 cm\(^{-1}\) with a 2 nm sampling interval resulted in total 1501 wavelengths. The NIRS were recorded at ambient temperature and a total of 72 spectra of samples were collected.

![Figure 1. Spectra acquisition using NIRFlex N500 Fiber Optic Solids Cell spectrometer](image)

2.1.4. Data analysis. Spectra collection was obtained using the NIRWare 1 Software which was then converted into Excel file for worksheet calculation and data plot. The raw data obtained from the NIR instrument were in reflectance values which were then imported to the Unscrambler X version 10.5.1 (CAMO Software AS, Oslo, Norway) for multivariate analysis, including spectra pre-processed
(Standard Normal Variate (SNV), 1st and 2nd Savitzky-Golay Derivative, 2 Polynomial Order, 25 Smoothing Points, and combination of Derivative and SNV), PCA analysis, and SIMCA.

3. Results and Discussion

3.1. Spectral data

Figure 2 shows the average spectra of flour samples in spectra range of 4000 – 10,000 cm$^{-1}$. Three pre-processed methods (i.e. Standard Normal Variate (SNV), 1st and 2nd Savitzky-Golay Derivative (with a second-order polynomial and a 25 smoothing points), as well as a combination of 2nd Savitzky-Golay Derivative (with a second-order polynomial and a 25 smoothing points) and SNV) were applied to spectra data. Based on the analysis (data were not shown), the selected pre-processed was the combination of 2nd Savitzky-Golay Derivative (with a second-order polynomial and a 25 smoothing points) and SNV (Figure 3).

![Figure 2](image_url)

Figure 2. Average spectra (log (1/R)) of mocaf, arrowroot, bread fruit, banana, sweet potato, and taro flour

Samples used in this study are all characterized by complex hydrogen bonding interactions between sugars and protein. The average spectra was dominated by several peaks located at 4545 – 6250 cm$^{-1}$ which correspond with 2nd overtones for N-H bonds, C-H stretching, C-O stretch combination, and the C-O stretch or N-H stretching vibrations which relate to protein [16].

![Figure 3](image_url)

Figure 3. Average pre-processed spectra of flour samples of mocaf, arrowroot, bread fruit, banana, sweet potato, and taro in the spectral region of 4000 – 10,000 cm$^{-1}$
3.2. **Classification using Principle Component Analysis (PCA)**

Table 1 shows the percentage variance and cumulative percentage for the 7 principal components (PC). The first 4 PC accounted for 91% of the total variation. The pair-wise plot of PC1 and PC2 scores which explains 72% of the total variation, shows an overlap between taro and mocaf flour. However, the other flours are closely grouped. By using PC1 (38%) and PC3 (16%) scores for the flour samples shown in Figure 4, clearly indicates six clusters of flour samples. PC1 separates sweet potato, breadfruit, banana, taro / mocaf, and arrowroot. PC2 is able to separate mocaf, banana, and sweetpotato / breadfruit / taro / banana. The mocaf flour in Figure 4 is clearly separated from other flours due to its lowest protein content [17].

| Principle components | Percentage of variance | Cumulative percent |
|----------------------|------------------------|--------------------|
| PC1                  | 38                     | 38                 |
| PC2                  | 34                     | 72                 |
| PC3                  | 16                     | 88                 |
| PC4                  | 3                      | 91                 |
| PC5                  | 2                      | 93                 |
| PC6                  | 1                      | 94                 |
| PC7                  | 1                      | 95                 |

It can be observed that several of the flours are characterized by either negative score values for the PC1 or negative scores for the PC3. The loadings of corresponding PCs underlined important spectral regions associated with the classification observed. The loadings for the PC1 and PC2 are dominated by peaks associated with N–H absorptions in the regions of 5347- 5988 cm⁻¹ which are connected with protein. Those spectral regions accounts for the discrimination between the various flours.

![Figure 4. PC1 ×PC3 score plot for the overall set of calibration samples. The explained variance by each PC was indicated in parenthesis.](image)

3.3. **SIMCA classification**

The training set for each type of flours was based on 36 spectral measurements. The model validation was based on full cross validation. The classification set was built using six spectra for each type of flours. For five flour samples (except purple sweet potato), the samples from the set are classified...
correctly. Figure 5 shows an example of class separation by using the Coomans plot at the 75% confidence level shows the orthogonal distances of samples from the classification set from the models of arrowroot flour versus modified cassava flour. Similar results are obtained for the other Coomans plots.

![Figure 5. Example of Coomans plot of arrowroot flour versus modified cassava flour](image)

**Conclusions**

The NIR spectrum of flour used in this study are rich in chemical information; therefore, by applying multivariate technique it can be used to differentiate between samples. The SIMCA can classify 100% success rate for mocaf, banana, arrowroot, bread fruit, and taro, and 67% for purple sweet potato. NIR spectroscopy coupled with proper multivariate analysis is a valuable tool for identification of samples.

**Acknowledgement**

This research was funded by Universitas Gadjah Mada under Recognisi Tugas Akhir program 2019 No: 2129/UN1/DITLIT/DIT-LIT/LT/2019. Great appreciation is delivered to Nafisatul Maghfiroh, Win Pribadi, and Tafip Hariyanto (UKSW Salatiga) for their support and assistance in finishing this work.

**References**

[1] Lin M A, Wu M, Lu S, Lin J, Lin M A, Wu M and Lin J 2010 Glycemic index, glycemic load and insulinenic index of Chinese starchy foods 16 4973–9

[2] Ebere R A, Imungi J K and Kimani V N 2017 Glycemic Indices of Cassava and Sweet Potatoes Consumed in Western Kenya 63 7–12

[3] Onyango C 2016 Starch and modified starch in bread making: A review African J. Food Sci. 10 344–51

[4] Aryee F N A, Oduro I, Ellis W O and Afuakwa J J 2006 The physicochemical properties of flour samples from the roots of 31 varieties of cassava Food Control 17 916–22

[5] Wang S, Nie S and Zhu F 2016 Chemical constituents and health effects of sweet potato Food Res. Int. 89 90–116

[6] Huang A S, Titchenal C A and Meilleur B A 2000 Nutrient Composition of Taro Corms and Breadfruit J. Food Compos. Anal. 13 859–64

[7] Chen J, Zhu S and Zhao G 2017 Rapid determination of total protein and wet gluten in commercial wheat flour using siSvr-NIR Food Chem. 221 1939–46

[8] Sohn M, Barton F E, McClung A M and Champagne E T 2004 Near-Infrared Spectroscopy for Determination of Protein and Amylose in Rice Flour Through Use of Derivatives Cereal Chem. 81 341–4
[9] Shimizu N, Jyunji Katsura, Takashi Yanagisawa, Bunji Tezuka, Yasuyuki Maruyama S I, Colin G. Eddison, Ian A. Cowe R P W, Blakeney A B, Toshinod Kimura S Y and Hiroshi Okadome, Hidechika Toyoshima K O 1998 Evaluating techniques for rice grain quality using near infrared transmission spectroscopy J. Near Infrared Spectrosc. 6 111–6

[10] Rothkaehl J 1998 The near infrared-instruments in grain and milling industry in Poland - The current situation and the future J. Near Infrared Spectrosc. 6 97–9

[11] Miyamoto K 1998 Classification of high acid fruits by partial least squares using the near infrared transmittance spectra of intact satsuma mandarins 271 267–71

[12] Naila A, Flint S H, Sulaiman A Z, Ajit A and Weeds Z 2018 Classical and novel approaches to the analysis of honey and detection of adulterants Food Control 90 152–65

[13] El-Abassy R M, von der Kammer B and Materny A 2015 UV Raman spectroscopy for the characterization of strongly fluorescing beverages LWT - Food Sci. Technol. 64 56–60

[14] Huang M, Kim M S, Delwiche S R, Chao K, Qin J, Mo C, Esquerre C and Zhu Q 2016 Quantitative analysis of melamine in milk powders using near-infrared hyperspectral imaging and band ratio J. Food Eng. 181 10–9

[15] Lohumi S, Lee S, Lee W H, Kim M S, Mo C, Bae H and Cho B K 2014 Detection of starch adulteration in onion powder by FT-NIR and FT-IR spectroscopy J. Agric. Food Chem. 62 9246–51

[16] Magwaza L S, Naidoo S I M, Laurie S M, Laing M D and Shimelis H 2016 Development of NIRS models for rapid quantification of protein content in sweetpotato [Ipomoea batatas (L) LAM] LWT-Food Sci. Technol. 72 63–70

[17] Yuliyanda I, Masithoh R E, Khuriyati N and Saputro A D 2019 Classification of flour from various crops based on protein contents using near infra-red spectroscopy and principle component analysis (Universitas Gadjah Mada)