Determination of Total Flavonoid Content in Medicinal Plant Leaves Powder Using Infrared Spectroscopy and Chemometrics

Lestyo Wulandari[*], Bayu Dwi Permana, and Nia Kristiningrum

Pharmaceutical Analysis and Chemometrics Research Group, Faculty of Pharmacy, University of Jember, Jl. Kalimantan 37, Jember 68121, East Java, Indonesia

Abstract: *Flavonoid is phenolic compounds consisting of fifteen carbon atoms and is commonly found in plants. Infrared (IR) spectroscopy combined with chemometrics, has been developed for a simple analysis of flavonoid in the medicinal plant leaves powder. IR spectra of selected medicinal plant powder were correlated with flavonoid content using chemometrics. The chemometric methods used for calibration analysis were Partial Least Square (PLS), Principal Component Regression (PCR), and Support Vector Regression (SVR). After the calibration model was formed, it was then validated using Leave-One-Out-Cross-Validation (LOOCV) and 2-fold cross-validation. In this study, the PLS of the Near-infrared (NIR) calibration model showed the best calibration with R-Square and RMSEC values of 0.9676524 and 0.0978202, respectively. The LOOCV of PLS of the NIR calibration model has the R-square and RMSE values of 0.9850164 and 0.067663, respectively. The 2-fold cross-validation gave the R-square and RMSE values of 0.9857071 and 0.2104665, respectively. PLS of the NIR calibration model was further used to predict unknown flavonoid content in commercial samples. The significance of flavonoid content that has been measured by NIR and UV-Vis spectrophotometry was evaluated with paired samples T-test. The flavonoid content that has been measured with both methods gave no significant difference.*

Keywords: *medicinal plant leaves powder; total flavonoid content; NIR; FTIR; chemometric*

■ INTRODUCTION

Traditional medicine is a mixture of ingredients from plants, animals, and minerals that have been used for generations in accordance with the prevailing norms in society [1]. The use of traditional medicines in the past decade has increased along with the development of lifestyle called "back to nature" and rising prices of modern medicines. Indonesia, according to Torri [2], is a country with a biodiversity rank 3rd in the world where there are around 30,000 species out of a total of 40,000 species of plants. However, only 25% or about 7,500 plant species have been known or examined for their potential use as medicinal plants. This shows that there are still many species of plants in Indonesia that have no known potential as a medicinal plant, so that the research about active chemical compounds in plants needs to be done.

Parts of the plant that are often used for medicines and research materials are the leaves. This is related to some advantages of leaves, such as ease of use, more productivity, and readily available than other parts of the plant [3]. One of the active chemical compounds that are often found in plants, especially in parts of the leaves is flavonoid [4]. Flavonoid is phenolic compounds consisting of 15 carbon atoms and is commonly found in plants [5]. Flavonoid has several pharmacological activities, such as antioxidants, hepatoprotective, antibacterial, anti-inflammatory, anticancer, antiviral and vasodilator [4].

Several analytical techniques have been developed for determining flavonoid content such as UV spectrometry, mass spectrometry, gas chromatography (GC), thin-layer chromatography, high-performance liquid chromatography (HPLC) and capillary

electrophoresis [6]. These methods can measure accurately, but the process is too long and requiring various reagents or solvents that are quite expensive [6]. Based on this case, a simple, selective, and eco-friendly method is required.

Infrared (IR) spectroscopy method is an analytical method that has several advantages, such as non-destructive (does not damage the sample), does not cause pollution, does not require chemical reagents (without solvents) and can analyze at high speed [7]. On the other side, infrared spectroscopy has the disadvantage that the data formed from the results of the analysis are quite complicated and with overlapping spectra [8], therefore multivariate chemometrics analysis is needed to analyze data from the infrared spectra. The chemometric calibration model is a statistical and mathematical analysis technique for processing, evaluating, and interpreting

relevant information from (Near-infrared) NIR and (Fourier Transform-infrared) FTIR spectra data. This method can be used to analyze NIR and FTIR spectrum data, which have overlapping absorption bands [9]. The functions of chemometrics are to find good statistical correlations between spectral data and chemical parameter data for the model development process. Chemometric methods are often applied to a condition where no theory can solve or describe a problem that uses many variables (multivariate). The purpose of forming a chemometric calibration model is to describe a system that has a hidden relationship between available data and the information generated [8]. The quantitative multivariate models that are often used are Partial Least Square (PLS), Principal Component Regression (PCR), and Support Vector Regression (SVR) [10].

The objective of this research is to develop a simple, rapid and validated model of infrared spectroscopy at different altitudes by NIR and FTIR spectroscopies with chemometrics for the determination of the flavonoid content. Furthermore, the selected infrared spectroscopy calibration model with chemometrics methods was applied for determining the flavonoid content in commercial samples.

Table 1. Plant species used in this research

No	Scientific name (Species)	Code
1.	*Ocimum basillicum*	A
2.	*Cosmos caudatus*	B
3.	*Coffea canephora*	C
4.	*Nepheliumlappaceum*	D
5.	*Persea americana*	E
6.	*Averrhoa bilimbi*	F
7.	*Pluchea indica*	G
8.	*Magnolia liliifera*	H
9.	*Moringa oleifera*	I
10.	*Morinda citrifolia*	J
11.	*Syzgium polyanthum*	K
12.	*Psidium guajava*	L
13.	*Clidemia hirta*	M
14.	*Sauropus androgynus*	N
15.	*Lunasia amara*	O
16.	*Aleurites moluccana*	P
17.	*Artocarpus heterophyllus*	Q
18.	*Diplazium esculentum*	R
19.	*Piper betle*	S
20.	*Annona muricata*	T
21.	*Syzygium aqueum*	U
22.	*Leucaena leucocephala*	V
23.	*Chrysophyllum cainito*	W
24.	*Pandanus sp*	X
25.	*Carica papaya*	Y

■ EXPERIMENTAL SECTION

Materials

The materials used in this study were twenty-five (25) samples of medicinal plant leaves with varieties of flavonoid content taken from the residential areas and plantation areas in Jember city, East Java, Indonesia (Table 1). Reagents used were ethanol 96% (Brataco, Indonesia), quercetin (Sigma-Aldrich), 10% aluminum chloride, potassium acetate (Merck), and filter paper. Five commercial samples were purchased from a pharmacy department store in Jember city (Chinese Teak®, Meniran®, Sambiloto®, Gotu Kola®, and Soursop®).

Instrumentation

The instruments used were NIR spectrometer (Brimrose Corporation Luminar 3070) Brimrose Software (Brimrose corp.), The Unscrambler X 10.2 software (CAMO), computer (Intel Core 2 Duo), FTIR

spectrometer (Bruker Alpha) and UV-Vis spectrophotometer (Hitachi U 1800), moisture analyzer (PMB 53), grinding Machine (Philips HR2116), stainless steel sieves 100 mesh (Advantech), analytical balance (Sartorius), and oven (Memmert).

Procedure

Preparation and testing of the water percentage content of the leaves powder

The leaves that had been collected were then pulverized using a grinding machine and sieved using a 100 mesh sieve. The concentration of water content was tested using a moisture analyzer until the water content was less than 10% [11]. If the water content of the sample was greater than 10%, the powder was dried again in an oven at 50 °C until the water content of the sample was less than 10%.

Analysis using NIR and FTIR spectrophotometers

All samples were scanned on both infrared spectroscopy instruments (NIR and FTIR). Each sample was examined five replications by NIR spectrometer with five shots for each replication and three replications by the FTIR spectrometer [12]. The NIR and FTIR spectra data were obtained through Acquire Brimrose (NIR) and OPUS (FTIR) software, and each spectra data were coded.

Determination of the total flavonoid content

In this research, total flavonoid content was determined by UV-Vis spectroscopy as a reference method. This method used aluminum chloride as the reagent and quercetin as the reference compound [13]. The sample was prepared by mixing 0.5 mL of 10 mg/mL sample extract in ethanol with 3 mL of ethanol, 0.2 mL of $AlCl_3$ 10%, 0.2 mL of 1 M potassium acetate and then diluted to 10 mL in distilled water. After incubation at room temperature for 30 min, the absorbance of the mixture solution was measured at 428 nm by using a UV-Vis spectrophotometer. Various standard solutions of quercetin (2.0, 4.0, 6.0, 8.0, 10.0, 12.0 and 15.0 μg/mL) were also prepared.

Determination and validation of the calibration model

Spectral data of the samples that had been scanned using NIR and FTIR spectrophotometers were analyzed using the Unscrambler X 10.2 (Camo Analytics, Norway) software. The spectral data of the training sets were correlated with the flavonoid concentration of the samples to form the calibration model. The calibration models used were PLS (partial least square), PCR (principal component regression), and SVR (support vector regression).

PLS model is a multivariate projection method for modeling a relationship between dependent variables (Y) and independent variables (X). The principle of PLS is to find the component in the input matrix (X) that describe as much as possible the relevant variations in the input variables and at the same time have a maximal correlation with the target value in Y, given less weight to the variations that are irrelevant or noisy [14]. Therefore, PLS models, both X and Y, simultaneously to find the latent variables in X that will predict the latent variables in the Y model. The PCR model has a principle like PLS, but the difference is that PCR can eliminate several predictors by using several major components [15]. The SVR model is a result of the development of a regression method that can be used to predict the regression method in a statistical approach. The SVR principle is based on the calculation of linear regression functions in high-dimensional spaces where the input data is mapped through nonlinear functions [16].

The concentration of flavonoid of samples determined using UV-Vis spectroscopy as a response (variable y) and absorbance value of infrared spectra as a predictor (variable x). The assessment criteria in forming a calibration model were the value of R^2, RMSEC (root mean square error of calibration), and RMSECV (root mean square error of cross-validation). The calibration model was stated to be good and has the best prediction if the correlation value R^2 is greater (greater than 0.91 or close to 1), and the RMSEC and RMSECV values were getting smaller [15].

The calibration model selected was then validated using a Leave-One-Out-Cross-Validation (LOOCV), and 2-Fold-Cross-Validation (2-FCV) approaches. LOOCV was evaluated by omitting a data point, or a sample from the training data set one at a time. Then, the remaining data sets or samples were used to build a

new calibration model, which was then used to predict the omitted data set or sample. The second validation method was 2-fold cross-validation. In this research, validation with 2-fold cross-validation used 5 independent samples (test sets) with known concentrations. The calibration model formed were used to predict the concentration of flavonoid in the test set samples [17].

Applications of the selected and validated model for commercial samples

The selected and validated calibration models were applied for the determination of the total flavonoid content in the commercial samples. A spectrum of commercial samples was scanned then was analyzed by a chemometric calibration model in order to predict the concentration of the total flavonoid content [12].

Analysis of data

The results of the determination of the total flavonoid content in the commercial samples by infrared spectroscopy and UV-Vis spectrophotometry methods were compared using the Two Paired Samples t-Test with SPSS Ver. 22 (IBM, USA). Data analysis was carried out at a confidence level of 95% with significance or Sig. (2-tailed) > 0.05 [18-19].

■ RESULTS AND DISCUSSION

Preparation and Testing of Water % Content

Each of the samples was given a unique code, and its moisture content of < 10% was presented in Table 2. The

water content of sample powder has to prepared below 10% in order to reduce the interfere of spectra measurements of the sample and to stop the process enzymatic reactions of microbial that can reduce the quality of sample powder [12]. The results of NIR and FTIR spectra data are shown in Fig. 1 and the spectral data of all samples were used as predictors on the calibration models.

Total Flavonoid Content in the Sample

Total flavonoid content in the sample was determined in mg quercetin equivalent (QE). Quercetin is used as a standard (comparison) because it belongs to the largest groups of flavonol that are often found in plants, where its amount is between 60–75% of flavonoids. In addition, quercetin also has keto groups on C-4 atoms and hydroxyl groups on C-3 or C-5 atoms, which can react to form acidic complexes with $AlCl_3$ [13]. The results of the total flavonoid content are as shown in Table 3. The range of total flavonoids is 3–30 mg QE/g powder. The samples were then divided into two categories, namely training sets (used in forming the calibration models) and test sets (used for validation on selected calibration models).

NIR and FTIR Calibration Models

The calibration model was formed by correlating NIR and FTIR spectra of the samples with the total flavonoid content that was determined by spectroscopy

Table 2. Code name and % water content of the sample

No	Code	% of water content ± SD	No	Code	% of water content ± SD
1.	A	4.633% ± 0.0764	14.	N	5.250% ± 0.1000
2.	B	5.650% ± 0.0500	15.	O	3.500% ± 0.0866
3.	C	4.767% ± 0.1041	16.	P	4.200% ± 0.0500
4.	D	5.217% ± 0.0289	17.	Q	3.233% ± 0.0764
5.	E	4.217% ± 0.0764	18.	R	4.750% ± 0.2000
6.	F	4.300% ± 0.0500	19.	S	5.183% ± 0.1528
7.	G	4.150% ± 0.0500	20.	T	4.750% ± 0.1323
8.	H	4.250% ± 0.1000	21.	U	3.133% ± 0.0764
9.	I	3.883% ± 0.0764	22.	V	4.500% ± 0.0500
10.	J	4.783% ± 0.0764	23.	W	4.233% ± 0.1041
11.	K	4.483% ± 0.1258	24.	X	4.500% ± 0.1000
12.	L	2.867% ± 0.0764	25.	Y	4.133% ± 0.0764
13.	M	3.567% ± 0.0289			
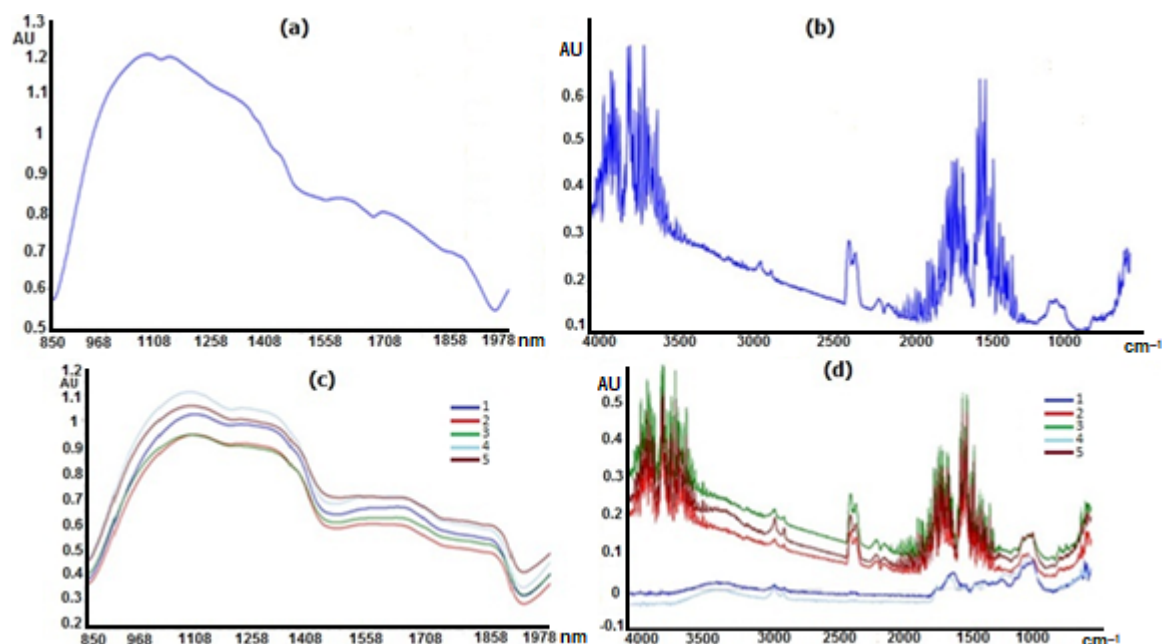
Fig 1. (a) NIR spectra of quercetin (b) FTIR spectra of quercetin (c) NIR spectra of samples (d) FTIR spectra of samples, 1) *Ocimum basillicum*, 2) *Nephelium lappaceum*, 3) *Coffea canephora*, 4) *Cosmos caudatus*, 5) *Persea americana*

Table 3. The results of the total flavonoid content in the training set and test sample

Code	Category sample	Average mg QE/g ± SD	Code	Category sample	Average mg QE/g ± SD
A	*Training Set*	13.22 ± 0.0273	N	*Training Set*	15.11 ± 0.0479
B	*Training Set*	29.93 ± 0.0136	O	*Training Set*	10.39 ± 0.0118
C	*Training Set*	8.108 ± 0.0236	P	*Training Set*	7.361 ± 0.0136
D	*Training Set*	9.439 ± 0.0295	Q	*Training Set*	7.985 ± 0.0277
E	*Training Set*	14.93 ± 0.0236	R	*Training Set*	9.017 ± 0.0518
F	*Training Set*	13.24 ± 0.0174	S	*Training Set*	11.46 ± 0.0236
G	*Training Set*	12.49 ± 0.0177	T	*Training Set*	14.47 ± 0.0169
H	*Training Set*	9.176 ± 0.0192	U	*Test Set*	8.139 ± 0.0236
I	*Training Set*	16.35 ± 0.0320	V	*Test Set*	26.30 ± 0.0459
J	*Training Set*	8.688 ± 0.0086	W	*Test Set*	3.726 ± 0.0115
K	*Training Set*	10.57 ± 0.0136	X	*Test Set*	4.048 ± 0.0216
L	*Training Set*	12.91 ± 0.0107	Y	*Test Set*	17.92 ± 0.0123
M	*Training Set*	5.350 ± 0.0136			

UV-Vis as the reference method. In order to create a calibration model, 362 data were used as the training set in the NIR, while 60 data were used as the training set in the FTIR. The calibration models used were PLS, PCR, and SVR. The results of the calibration model are as shown in Table 4.

In this research, the calibration model selected and used in determining the total flavonoid content was the PLS model in the NIR. The PLS model in the NIR was chosen because it has the R-square value that is close to 1

(0.9676524) and the smallest RMSEC value (0.0978202). A calibration model is considered optimum or has a good correlation if it has an R-square value close to 1 and the smallest RMSEC value [20].

The selected PLS model was then validated with Leave-One-Out-Cross-Validation (LOOCV) and 2-Fold-Cross-Validation (2-FCV). The cross-validation was carried out for determining the accuracy and predictive ability of the models that have been formed [21]. The results of the LOOCV of PLS for the NIR calibration

Table 4. The results of PLS, PCR and SVR calibration models from NIR and FTIR spectra

Spectra	Model	R-*square* Calibration	R-*square* Validation	RMSEC	RMSECV
NIR	PLS	0.9676524	0.9628204	0.0978202	0.1050252
	PCR	0.9592377	0.9526183	0.1098087	0.1189304
	SVR	0.5125262	0.4877062	0.4065073	0.4067137
FTIR	PLS	0.9240598	0.7372873	0.1384766	0.2625193
	PCR	0.8064864	0.6212984	0.2626473	0.3743665
	SVR	0.9136168	0.5951188	0.1674283	0.3239591

model show the R-square and RMSE values are 0.9850164 and 0.067663, respectively. The 2-fold cross-validation results show the R-square and RMSE values are 0.9857071 and 0.2104665, respectively (Fig. 2). Based on the results of LOOCV and 2-fold cross-validation, it can be concluded that the PLS calibration model formed has good predictability and accuracy. The model can be applied to commercial samples.

Selected and Validated Model Applications in the Commercial Samples

The commercial samples were scanned using a NIR spectrophotometer five times with five replicates each time, and as a result, 125 data were obtained. All of the collected data were analyzed by being included in the Unscrambler X 10.2 software to determine the total flavonoid content of the sample. Furthermore, all of the data of the total flavonoid content of the samples (mg QE/g powder) obtained using NIR spectroscopy were compared with the results of the sample content obtained using UV-Vis spectrophotometry and the SPSS software. The results of the determination of the total flavonoid content in the commercial samples are as shown in Table 5.
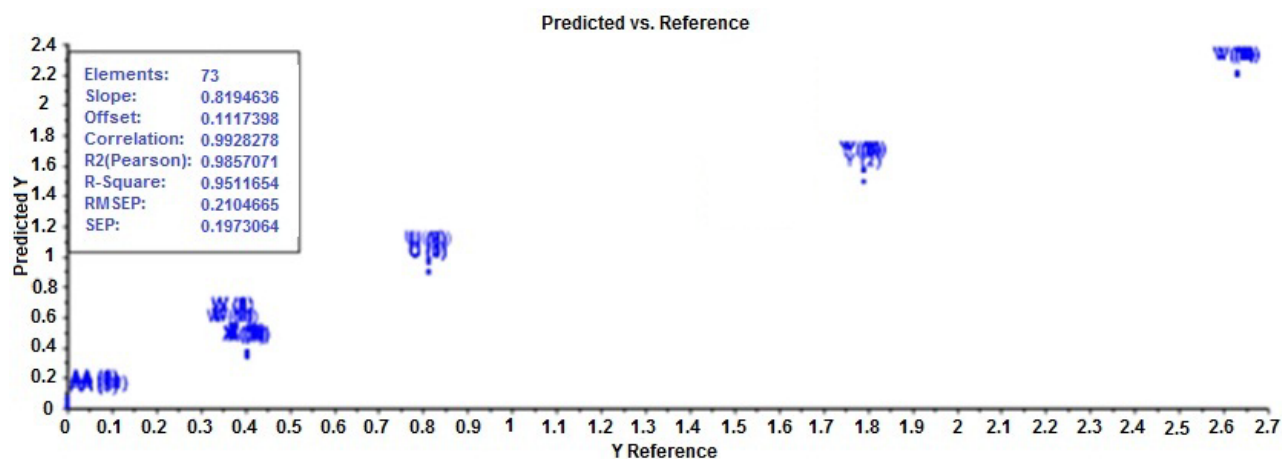
Fig 2. The result of a 2-fold cross-validation PLS model

Table 5. The results of the determination of the total flavonoid content in the commercial samples

Real Sample	Average of total flavonoid content (mg QE/g) ± SD	
	NIR Spectroscopy	UV-Vis Spectrophotometry
Chinese Teak®	17.95 ± 0.4119	18.12 ± 0.0136
Meniran®	12.14 ± 0.0503	12.26 ± 0.0136
Sambiloto®	12.48 ± 0.0964	12.58 ± 0.0307
Gotu Kola®	11.25 ± 0.1350	11.09 ± 0.0307
Soursop®	13.98 ± 0.0973	14.15 ± 0.0545

The results of total flavonoid content in the commercial samples using NIR spectroscopy and UV-Vis spectrophotometry were compared using the "Paired Sample T-test". The purpose was to find out whether there was a significant difference between the total flavonoid content in the commercial samples determined using the NIR spectroscopy over UV-Vis spectrophotometry. In the paired sample T-test, decision forming is based on its significance value, if the significance (Asymp. Sig) < 0.05, then Ho (no significant difference) is rejected and Ha (there are significant differences) accepted. Conversely, if the significance (Asymp. Sig) > 0.05, then Ho is accepted, and Ha is rejected. The results of the paired sample T-test showed that there were no significant differences in the determination of the total flavonoid content in the commercial samples with the two methods, where the p-value obtained was 0.249 with a 95% confidence level.

■ CONCLUSION

The NIR spectroscopy, combined with multivariate calibrations methods, can be used to determine the total flavonoid content in medicinal plant leaf powder. The selected calibration model was PLS (Partial Least Square) with R-square, and RMSE calibration values were 0.9676524 and 0.0978202, respectively. The suggested method is simple, selective, validated, and eco-friendly.

■ REFERENCES

[1] Yuan, H., Ma, Q., Ye, L., and Piao, G., 2016, The traditional medicine and modern medicine from natural products, *Molecules*, 21 (5), 559.

[2] Torri, M.C., 2013, Traditional jamu versus industrial jamu: Perceptions and beliefs of consumers in the city of Yogyakarta: What future for traditional herbal medicine in urban Indonesia?, *IJESB*, 19 (1), 1–20.

[3] Yusro, F., Mariani, F., Diba, F., and Ohtani, K., 2014, Inventory of medicinal plants for fever used by four Dayak sub ethnic in West Kalimantan, Indonesia, *Kuroshio Sci.*, 8 (1), 33–38.

[4] Kumar, S., and Pandey. A.K., 2013, Chemistry and biological activities of flavonoids: An overview, *Sci. World J.*, 2013, 162750.

[5] Brodowska, K.M., 2017, Natural flavonoids: Classification, potential role, and application of flavonoid analogues, *Eur. J. Biol. Res.*, 7 (2), 108–123.

[6] Siddiqui, M.R., AlOthman, Z.A., and Rahman, N., 2017, Analytical techniques in pharmaceutical analysis: A review, *Arabian J. Chem.*, 10 (Suppl. 1), S1409–S1421.

[7] Haas, J., and Mizaikoff, B., 2016, Advances in mid-infrared spectroscopy for chemical analysis, *Annu. Rev. Anal. Chem.*, 9, 45–68.

[8] Gad, H.A., El-Ahmady, S.H., Abou-Shoer, M.I., and Al-Azizi, M.M., 2013, Application of chemometrics in authentication of herbal medicines: A review, *Phytochem. Anal.*, 24 (1), 1–24.

[9] Ozaki, Y., 2012, Near-infrared spectroscopy-its versatility in analytical chemistry, *Anal. Sci.*, 28 (6), 545–563.

[10] Morellos, A., Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., and Mouazen, A.M., 2016, Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy, *Biosyst. Eng.*, 152, 104–116.

[11] Hunt Jr., E.R., Daughtry, C.S.T., and Li, L., 2016, Feasibility of estimating leaf water content using spectral indices from worldview-3's near-infrared and shortwave infrared bands, *Int. J. Remote Sens.*, 37 (2), 388–402.

[12] Wulandari, L., Retnaningtyas, Y., and Lukman, H., 2016, Analysis of flavonoid in medicinal plant extract using infrared spectroscopy and chemometrics, *J. Anal. Methods Chem.*, 2016, 4696803.

[13] Mathur, R., and Vijayvergia, R., 2017, Determination of total flavonoid and phenol content in *Mimusops elengi*, *Int. J. Pharm. Sci. Res.*, 8 (12), 5282–5285.

[14] Mantanus, J., 2012, New pharmaceutical applications involving near-infrared spectroscopy as a pat compliant process analyzer, *Dissertation*, Faculty of Medicine, University of Liege, Belgium.

[15] Lengkey, L.C.E.C., Budiastra, I.W., Seminar, K.B., and Purwoko, B.S., 2013, Determination of chemical properties in *Jatropha curcas* L. seed IP-3P by partial least-squares regression and near-infrared reflectance spectroscopy, *Int. J. Agric. Innov. Res.*, 2 (1), 41–48.

[16] Gu, B., Sheng, V.S., Wang, Z., Ho, D., Osman, S., and Li, S., 2015, Incremental learning for ν-support vector regression, *Neural Networks*, 67, 140–150.

[17] Ritz, M., Vaculikova, L., and Plevova, E., 2011, Application of infrared spectroscopy and chemometric methods to identification of selected minerals, *Acta Geodyn. Geomater.*, 8 (1), 47–58.

[18] Vasishth, S., and Nicenboim, B., 2016, Statistical methods for linguistic research: Foundational ideas – Part I, *Lang. Linguist. Compass*, 10 (8), 349–369.

[19] Kumar, R., and Kumar, V., 2015, Discrimination of various paper types using diffuse reflectance ultraviolet–visible near-infrared (UV-Vis-NIR) spectroscopy: Forensic application to questioned documents, *Appl. Spectrosc.*, 69 (6), 714–720.

[20] Georgieva, M., Nebojan, I., Mihalev, K., Yoncheva, N., Kljusurić, J.G., and Kurtanjek, Z., 2013, Application of NIR spectroscopy and chemometrics in quality control of wild berry fruit extracts during storage, *Croatian J. Food Technol. Biotechnol. Nutr.*, 8 (3-4), 67–73.

[21] Cheng, H., Garrick, D.J., and Fernando, R.L., 2017, Efficient strategies for leave-one-out cross-validation for genomic best linear unbiased prediction, *J. Anim. Sci. Biotechnol.*, 8, 38.