The Measurement of Luwak Content in Coffee Blends Using UV-Visible Spectroscopy Combined with Support Vector Machine Regression (SVMR)

The Measurement of Luwak Content in Coffee Blends Using UV-Visible Spectroscopy Combined with Support Vector Machine Regression (SVMR)

D Suhandy[1,3] and M Yulia[2,3]

[1] Department of Agricultural Engineering, The University of Lampung, Indonesia.
[2] Department of Agricultural Technology, Lampung State Polytechnic, Indonesia.
[3] Spectroscopy Research Group (SRG), Laboratory of Bioprocess and Postharvest Engineering, Faculty of Agriculture, The University of Lampung, Bandar Lampung, Lampung, Indonesia.

E-mail: diding.sughandy@fp.unila.ac.id

Abstract. In this study, an investigation of using UV-Vis spectroscopy combined with support vector machine regression (SVMR) was evaluated for luwak coffee authentication. For this purpose, we provide 98 samples of coffee blends (mixed of luwak and non-luwak coffee). The range of luwak content in coffee blends was 50-100%. UV-Vis spectral data of the all samples were acquired using a UV-Vis spectrometer in transmittance mode. The coefficient of determination (R^2), root mean square error of calibration (RMSEC), and residual prediction deviation (RPD) were used to evaluate the accuracy of the developed SVMR calibration model. The results show that UV-Vis spectroscopy combined with SVMR method was a promising method to quantify luwak content in coffee blends with high R^2 and low RMSEC. The RPD value of more than three was also obtained.

1. Introduction

Partial least squares (PLS) regression is one of the most popular multivariate method applied for spectroscopic data analysis because of its simplicity to use, speed, relative good performance and easy accessibility [1-2]. Together with its variants and modifications, the PLS calibration model is the most widely used regression technique for constructing the relationship between spectral data and property of biological samples [3]. In PLS methodology, the spectrum– property relationship is assumed to be linear. This assumption is not always valid for all samples, and it is completely unacceptable for systems with strong intermolecular or intramolecular interactions [4].

To overcome this problem, several nonlinear methods have been proposed to develop regression model in spectroscopic data analysis including the extensive application of artificial neural networks (ANN) as the most effective and popular nonlinear methods [5]. However, ANN methods have several drawbacks such as the tendency to overfitting and the training time and computational resources: ANN training can take many hours, and even days, of CPU time even with modern computers [1]. Another alternative nonlinear method is based on support vector machines (SVM). SVMs are very interesting methods, simple in their theoretical background and very powerful in model and real world applications. Compared to ANN method, SVM methods have several advantages such as able to handle ill-posed problems, lead to global models that are often unique, computationally efficient nonlinear relationship

modeling through kernel mapping, and a superior generalization performance (lower error rates on test set) [6-7].

In the previous study, UV-Vis spectroscopy has been used together with PLS regression for detection and quantification of adulteration in luwak coffee with acceptable result [8]. However, due to variability of the resources of luwak coffee, a nonlinear relationship may be occurred and PLS regression may not be sufficient to handle it. For this reason, development a robust and global model based on nonlinear approach is needed. In the previous report, the regression models based on SVMs are sufficiently accurate and robust to be used for gasoline, biofuel, or diesel fuel analysis [1]. In the literature, recently some application of SVM regression for quantitative analysis in food quality were reported [2], [9-10]. However, the performance of SVMR, particularly on detection and quantification of adulteration in luwak coffee, is yet to be established. For this reason, the objective of the present study is to evaluate the application of SVM regression on the quantification of luwak content in coffee blend samples (mixed luwak and non-luwak coffee).

2. Materials and Methods

2.1. *Samples*

In this study, a number of 98 samples were prepared (1 gram weight for each samples). There are two types of samples: non-adulterated samples (consist of pure luwak coffee, 49 samples) and adulterated samples (blend of luwak and non-luwak coffee, 49 samples). The luwak content in blends has range 50-100% (w/w). An aqueous extraction procedure of the coffee samples was performed based on Suhandy and Yulia [9], [11]. For multivariate analysis, the samples were divided into two groups: calibration set (58 samples) and prediction set (40 samples).

2.2. *UV-Vis spectral data acquisition*

The UV-Vis spectral data of aqueous coffee samples were obtained in the range of 190-700 nm by using a UV-Vis spectrometer (Genesys™ 10S UV-Vis, Thermo Scientific, USA). This spectrometer was equipped with a quartz cell with optical path of 10 mm. The spectral acquisition was done at spectral resolution of 1 nm at a room temperature. The raw spectra (without any preprocessing) were used for further analysis.

2.3. *Support vector machine regression (SVMR)*

Support vector machines (SVM) were initially been developed by Vapnik and his co-workers [12-13] as a binary classification tool. SVM is one of machine learning method that has recently become popular and widely used and investigated because of its ability in prediction for both, classification and regression [14]. Theoretically, an SVM model is a representation of the training sample set as vectors in space mapped so that the samples from the separate categories are divided by a clear gap that is as wide as possible [1]. Unknown new samples from cross-validation or a test set are then mapped into that same space. Based on which side of the gap between classes they fall, they are predicted to belong to one category or another. These principles of SVM can easily be extended to regression tasks. More recently, SVM regression originally developed in the machine learning community [15-16]. For more detailed about theoretical background on SVMs for both classification and regression has been discussed in several reported studies [6], [12-13].

Practically, the central idea in SVR is that the models are built using a smaller, representative set of observations close to the classification or regression boundary. These are called support vectors (SVs). Using the epsilon-intensive loss function it is ensured that a global minimum is found and at the same time a generalization bound is optimized. Here, SVMR will be developed using The Unscrambler® 10.5 (trial version), a multivariate software from CAMO (Oslo, Norway). The SVR algorithm used within The Unscrambler® is based on code developed and released under a modified BSD license by Chih-Chung Chang and Chih-Jen Lin of the National Taiwan University [17].Two SVM regression types are available in The Unscrambler® 10.5 which are based on different means of minimizing the error

function of the regression: epsilon-SVMR and nu-SVMR. The parameter epsilon controls the width of the band where the cost of errors in the epsilon-intensive loss function is zero. The value of epsilon can affect the number of support vectors used to construct the regression function. The bigger the epsilon, the fewer support vectors are selected. In nu-SVMR, the nu value must be defined. Nu serves as the upper bound of the fraction of errors and is the lower bound for the fraction of support vectors.

In SVMR there is also a parameter C that determines the tradeoff between the model complexity (flatness) and the degree to which deviations larger than epsilon are tolerated. For example, if C is too large (infinity), then the objective is to minimize the empirical risk only, without regard to model complexity. In this software, the kernel type to be used can be chosen from the following four options: Linear, Polynomial, Radial basis function and Sigmoid. The linear kernel is simplest one and therefore less susceptible to overfitting and for this reason it will be used in this study.

3. Results and Discussion

3.1. *Spectral analysis*

Figure 1 shows the raw spectral data of coffee samples with different concentration of luwak coffee (50 and 100%) in the range 190-700 nm. It can be seen that the spectra highly overlap. High concentration (100%) of luwak content tends to have higher absorbance. High absorbance at 270 nm was observed and it may be related to absorbance of caffeine content in coffee samples [8]. The relatively high noise was also observed in the initial of spectral measurements at around 190 to 200 nm.

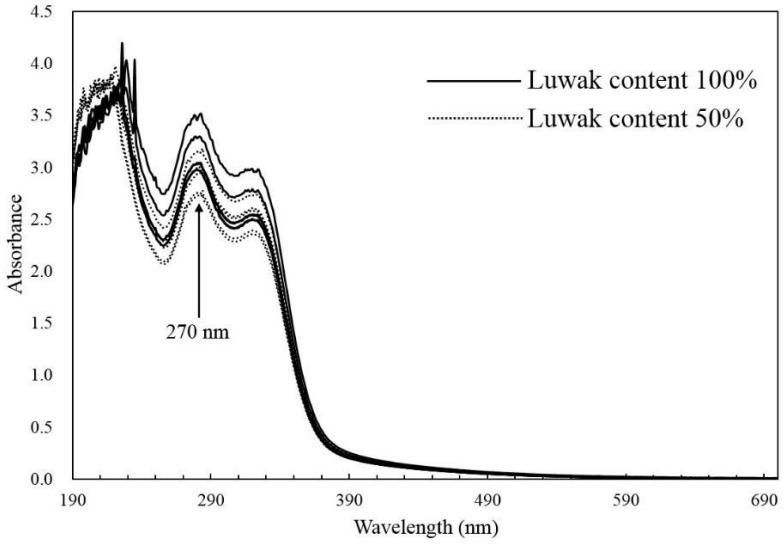

Figure 1. Absorbance spectra of coffee samples with low (50%) and high (100%) of luwak content in the range of 190-700 nm.

3.2. *Developing SVMR model*

Figure 2 shows the calibration model developed by using SVMR method using all wavelengths in the range of 190-700 nm. The SVMR model has high coefficient of determination (R^2) for calibration and validation and small differences between RMSEC and RMSECV. The SVR model was developed using the following parameters: epsilon-SVMR, kernel type = linear, C = 2 and Number of SVs = 34. The RPD of SVMR calibration model was quite high (4.31). This model was also comparable to previous report in luwak coffee quantification using PLS regression [8].

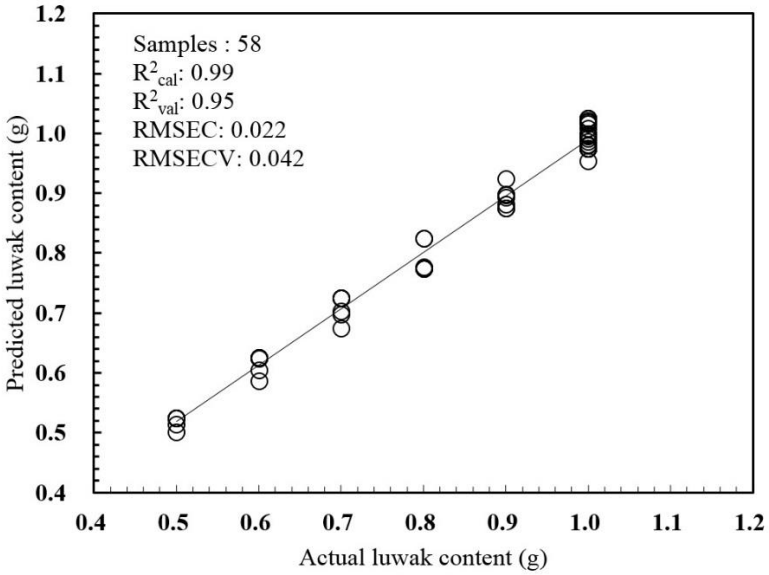

Figure 2. The result of SVMR model using SVM algorithm in the range of 190-700 nm.

3.3. *Evaluation of SVMR model*

To evaluate the performance of the developed SVMR model, a prediction using different samples was performed. The prediction of SVMR calibration model resulted in high coefficient of determination (R^2). Scatter plot between actual and predicted values is depicted in Figure 3. By a 95% confidence pair t-test, there were no significant differences between the actual luwak content and that predicted by UV-Vis spectroscopy. This result shows that a calibration model for simple and consistent determination of luwak content in luwak coffee blends using UV-Vis spectroscopy could be developed.

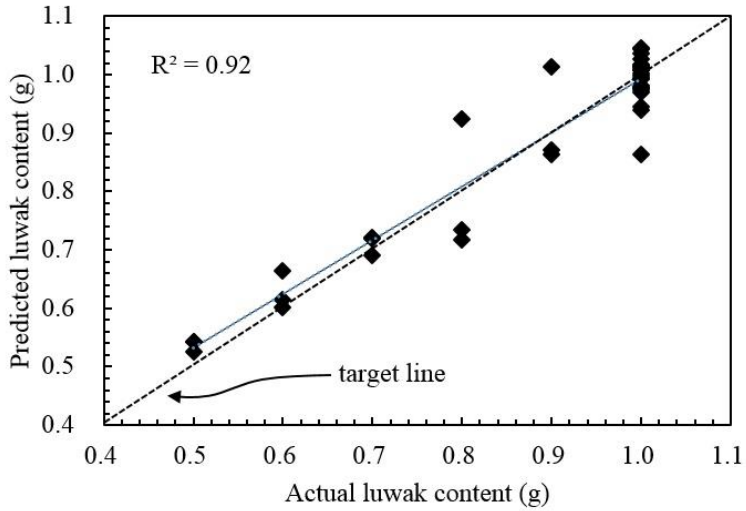

Figure 3. Scatter plot between actual and predicted luwak content in the prediction step.

4. Conclusion

UV-Vis spectroscopy, combined with support vector machines (SVM) regression model, is confirmed to be an interesting tool for simple and robust evaluation of luwak coffee authentication. The SVM regression algorithm gives good results for luwak content quantification in coffee blend samples (mixed luwak and non-luwak coffee). In conclusion, the combination of simple devices (UV-Vis spectrometer)

with non-linear modelling (SVM regression) may offer a very interesting and reliable tool for monitoring luwak coffee authentication.

Acknowledgement
The research was financially supported by Ministry of Research, Technology and Higher Education, Republic of Indonesia (Kemenristekdikti) under Research Grant PENELITIAN STRATEGIS NASIONAL INSTITUSI (PSNI) (Grant Number: 393/UN26.21/PN/2018). The authors also highly acknowledge Department of Agricultural Engineering, the University of Lampung for permission of using UV-Vis spectrometer and its supporting devices.

References
[1] Balabin R M and Lomakina E I 2011 *Analyst* **136**: 1703–12
[2] Malegori C, Marques E J N, de Freitas S T, Pimentel M F, Pasquini C and Casiraghi E 2017 *Talanta* **165**: 112–6
[3] Wold S, Sjostrom M and Eriksson L 2001 *Chemom. Intell. Lab. Syst*. **58**: 109–30
[4] Drain C M, Varotto A and Radivojevic I 2009 *Chem. Rev.* **109**: 1630–58
[5] Haykin S 1998 *Neural Networks: A Comprehensive Foundation* (Prentice Hall)
[6] Thissen U, Pepers M, Ustun B, Melssen W J and Buydens L M C 2004 *Chemom. Intell. Lab. Syst*. **73**: 169–79.
[7] Smola A J and Schölkopf B 2004 *Comput. Stat*. **14**: 199–222
[8] Suhand D and Yulia M 2017 *Int. J Food Sci* **2017***:*1–7
[9] Sanaeifar A, Bakhshipour A and de la Guardia M 2016 *Talanta* **148**: 54–61
[10] Xu S, Zhao Y, Wang M and Shi X 2017 *Catena* **157**: 12–23
[11] Suhandy D and Yulia M 2017 *Int. J Food Prop.* **20**: S331-9
[12] Bishop C M 2007 *Pattern Recognition and Machine Learning* (Springer)
[13] Vapnik V N 1995 *The Nature of Statistical Learning Theory* (New York: Springer Verlag)
[14] Ghasemi-Varnamkhasti M, Mohtasebi S S, Siadat M, Ahmadi H and Razavi S H 2015 *Eng. Agric. Environ. Food* **8**: 44–51
[15] Cortes C and Vapnik V 1995 *Mach. Learn.* **20**: 273–97
[16] Collobert S and Bengio S 2001 *J Mach. Learn. Res.* **1**: 143–60
[17] Chang C C and Lin C J 2011 *ACM Trans. Intell. Syst. Technol.* **2**:1–27