Performance of A Portable NIR Spectrometer to Distinguish Coffee Species Based on Qualitative Chemometric and Artificial Neural Network (ANN) Models

Agus Dharmawan[1], *Rudiati* Evi Masithoh[1] , and *Hanim* Zuhrotul Amanah[1]

[1] Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Gadjah Mada University, Indonesia

Abstract. A wide range of genetic cultivars of coffee and their characteristics determine consumer preference and increase industrial actors' awareness of production and marketing. The primary objective of this study is to develop a method to distinguish coffee species based on spectral characteristics acquired from a portable near-infrared spectrometer. The performance of this spectrometer in addressing classification problems is evaluated by the classification accuracy obtained from qualitative chemometrics, such as PCA and LDA, and artificial neural networks (ANNs) models. In this study, the instrument was successfully used and gained moderate accuracy for discriminating two coffee species, Arabica and Robusta, from Temanggung and Toraja. The accuracy was fair and achieved greater than 75%. Therefore, the instrument can be implemented as it provides simple, real-time, and in-situ analyses and can reach reliable results.

1 Introduction

Coffee is one of the leading plantation commodities in Indonesia, which has become a vital role in economic development. In 2021, the production of coffee reached 786.2 thousand tons, with a majority of 99.32% produced from smallholder plantations (1.26 ha). The export value of coffee was 387 thousand tons, with foreign exchange earnings of US$ 859 million [1]. Mostly, 97.17% of the exported coffee was Arabica WIB/Robusta OIB, not roasted, and not decaffeinated. Two species of coffee are widely cultivated and marketed in Indonesia, including Robusta (*Coffea canephora*) and Arabica (*Coffea arabica*). Most exported coffees are not roasted, not decaffeinated, wet-processed Arabica coffee beans and dry-processed Robusta coffee beans. Arabica coffee grows well at an altitude of 1000–2100 m above sea level with an air temperature of 18–22^{0}C and an annual rainfall of at least 1500 mm, whereas Robusta coffee grows well in a hot and humid climate at an altitude of 100–1000 m above sea level with an air temperature of 22–26^{0}C and an annual rainfall of at least 2000 mm [2], [3].

The quality of coffee varies as a result of differences in cultivar genetics and locations where they are grown, affecting their taste and making the price and popularity of coffee differ considerably [4]. Coffee quality is regularly examined based on chemical, sensory, and physical attributes. The chemical assessments of coffee beans are determined from the chemical composition contained [5]. The organoleptic properties relate to coffee aroma, flavor, sweetness, acidity, and overall taste. The physical characteristics include shape, thickness, weight, and color [6]. Coffee characteristics and consumer preferences have become significant *concerns in coffee industries, so industrial actors need to increase their awareness by confirming coffee species, varieties, and origins [4].

In distinguishing coffee species, the characteristics through visual aspects mostly play an important role [7]. The traditional recognition is carried out by experts/trained inspectors, is relatively high in cost, and the results may vary depending on the persons inspecting the food items. The physical assessments are difficult to identify and look similar because of their natural variability [8], [9]. Therefore, the accuracy of their inspections is frequently influenced by several factors, such as experts' moods, perceptions, and fatigue [10]. Spectroscopic methods in food analysis have now been developed and massively employed to address quantitative and qualitative issues. The quantitative issue is related to determining the chemical and compositional quality of foods. The qualitative issue is used to solve classification problems, including confirming geographic authenticity, distinguishing between species, the occurrence of adulteration, and the presence of defective beans [11]. The measurement procedure consists of (1) spectral data acquisitions, (2) data pre-processing to improve spectra, (3) building a classification model, and (4) evaluating the model's accuracy.

This study aimed to use a spectroscopic method, particularly in the NIR region, to distinguish coffee species. A portable pocket-sized near-infrared spectrometer (NIRS) was used to collect spectra from two species of coffee, Arabica and Robusta. Compared

to a benchtop NIR spectrometer, this portable spectrometer provides advantages such as non-destructive and in-situ analyses. The development has some critical factors, including measurement accuracy and high performance. But the implementation for industrial levels has increased interest, transforming large stationary analytical instruments into reliable-lightweight tools [12]. In this study, the discrimination between samples was carried out by qualitative chemometric models, including principle component analysis (PCA) and linear discriminant analysis (LDA), and a non-linear classification model, artificial neural network (ANN). Features in spectroscopic data is reduced using PCA and obtain lesser-dimensional data called principle components (PCs), which will become inputs to the classifiers, LDA and ANN.

Several studies have performed excellent classification results in distinguishing species of coffee using NIR spectroscopy. [13] used NIRS at a wavelength range of 1100–2500 nm to discriminate between arabica and robusta pure coffee varieties and blends of varied varietal composition. FTIRS was also used by [14] to discriminate coffee species by using PCA and classical discriminant analysis. [15] used FTIRS combined with KNN, SVM, PLS-DA, BP-ANN, and ELM to distinguish coffee species – Arabica and Robusta. The application of a portable spectrometer has only been found in a limited number of studies. [16] employed portable near-infrared spectroscopy to determine the quality of Arabica coffee by identifying and quantifying the adulterants with Robusta coffee (at different roasting levels) and adulterants with corn, peels, and sticks; the study combined two chemometric methods: partial least squares (PLS) and PCA. [17] employed handheld NIR spectroscopy combined with PLS-DA to evaluate directly cup profiles in roasted and ground coffee blends.

2 Methodology

2.1 Data Collection

The samples used in this study were green beans purchased from local markets in East Java (for Temanggung) and South Sulawesi (Toraja). The Arabica and Robusta samples were normal-shaped beans from full-wash and natural coffee processing, respectively, and had been cleaned from the remaining dry epidermis and endocarp/parchment. Spectral data were acquired using a pocket-sized connected micro-spectrometer (Consumer Physics, Inc. (SCiO), St. Cloud, Minnesota, USA). This spectrometer (Fig. 1) was integrated with user's mobile app to scan and real-time display and equipped with a back-end cloud to store the sample databases. This spectrometer worked in the NIR region, particularly 740 – 1070 nm, and generated 331 data points. A total of 2400 spectral data (2 species x 2 origin x 600 samples) were downloaded and stored in CSV format. The room temperature was maintained at 20±1 ^{0}C during spectral measurement. We only use raw spectra for further analyses.

Fig 1. Scheme of data collection using a portable NIR spectrometer

2.2 Qualitative Chemometric Models

PCA was used to compress spectroscopic data and reduce its variable from a high dimensionality to a lesser dimensionality. The transformed data was called principle components (PCs), which were linear combinations of the original data. The number of PCs obtained for further analyses was obtained from those close to 100% of the explained variance. The new data generated (PC scores) was scaled using Z-score normalization.

In this study, LDA would provide more class separability between the sample classes. In contrast to PCA, LDA required matrix *Y*, consisting of two class labels for two coffee species, Arabica and Robusta. The LDA results were evaluated by plotting LDA values against label classes of samples to visualize the samples' distinguishment. The model effectiveness was examined using *k*-fold cross-validation and confusion matrix; see sections 3.4 and 3.5.

2.3 Artificial Neural Networks Model

The ANN model consisted of an input layer, two hidden layers, and an output layer. The structure of the ANN model is depicted in Fig 2. The number of input nodes varied depending on the number of components generated in PCA. The five nodes were defined for both hidden layers. The output layer would classify two label classes, label 0: Robusta and 1: Arabica. The one-hot encoding technique converted the class label into a new categorical column, assigning a binary value of 1 or 0. The model effectiveness was evaluated using *k*-fold cross-validation. The confusion matrix, ROC curves and AUC values were also performed to visualize and check the ANN classifier's performance.

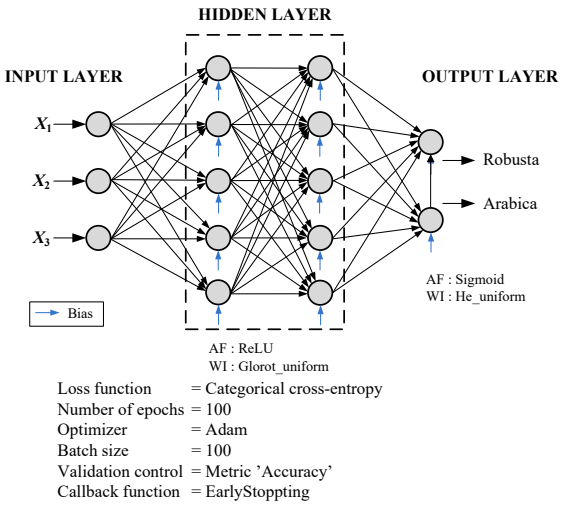

Fig 2. Structure of ANN classifier

2.4 Cross-Validation

The dataset was partitioned into a training set (randomly 2/3 of total samples) and a testing set (1/3 of random samples). *K*-fold cross-validation was used to evaluate the model's ability in a certain data to predict new data. This validation method used the split data from a training set. *k* referred to the number of folds (subsets) in the given data, the one-fold for testing and leaving the remaining $k - 1$ folds to train the model. The subsets were trained and validated in *k* iterations. The model's effectiveness was performed in mean+SD accuracy from all iterations.

2.5 Performance of Classification Models

A confusion matrix (Table 1) was used to evaluate the classification results of LDA and ANN. First, we must compute a set of predicted targets and compare them to the actual targets [18]. The predicted targets represented the values of the label classes as a result of the model, whereas the actual class represented the original values of the initial label classes [19]. The output TN (True Negative) showed the number of negative examples classified accurately, the TP (True Positive) indicated the number of positive examples classified accurately, the term FP (False Positive) value performed the number of actual negative examples classified as positive, and the FN (False Negative) was the number of actual positive examples classified as negative [20]. Positive samples were defined as classified Robusta beans, and negative samples represented classified Arabica beans.

Table 1. Scheme of two-class confusion matrix

		Predicted class	
		Positive	Negative
Actual class	Positive	TP	FN
	Negative	FP	TN

The evaluation of the model was calculated using performance metrics. Accuracy (AC) computed the percentage of correctly classified samples, see Eq 1, where *n* represented the total number of samples. Error (ER) was the percentage of wrongly assigned samples, and the formula is given in Eq 2. Recall (RE) was computed by the ratio of correctly classified samples to all positive samples, see Eq 3. Precision (PR) was the ratio of correctly classified samples as positive to all positively classified samples, as given in Eq 4. Specificity (SP) represented the ratio of incorrectly classified samples to all negative samples, see Eq 5. F1-score (Eq 6) was a weighted harmonic mean of PR and RE [21][22]. A good classifier was able to generate accuracy ideally close to 100% and error nearest to 0%. The specificity, precision, recall, and F1-Score were also declared 'good' if they were close to 100% or 1.00.

$$AC = \frac{TP + TN}{n} \tag{1}$$

$$ER = \frac{FP + FN}{n} \tag{2}$$

$$RE = \frac{TP}{TP + FN} \tag{3}$$

$$PR = \frac{TP}{TP + FP} \tag{4}$$

$$SP = \frac{TN}{TN + FP} \tag{5}$$

$$FS = \frac{2 \cdot R \cdot P}{R + P} \tag{6}$$

3 Experimental Results and Discussion

3.1 Spectral Profiles of Arabica and Robusta Coffee

A portable NIR spectrometer used in this study worked in the range of 740–1070 nm; the averaged spectra acquired are shown in Figure 3. In the NIR region, molecular vibration occurs when light strikes the samples and exhibits absorption bands in the molecules. Thus, in this region, we can also use it for chemical and compositional analyses of food products. The main absorption bands occurred at 900–1000 nm for the 3rd overtone of CH, CH_2, and CH_3 groups corresponding to phenolic compounds and caffeine content [17], [23][16]. The main difference between Arabica and Robusta coffee is their taste; Arabica coffee is more acidic, whereas caffeine carries a bitter taste in Robusta [24]. Robusta beans contain higher levels of caffeine than Arabica beans. Bitterness in Robusta is highly related to caffeine content [25]. Caffeine in the NIR Region corresponds to CH stretching [26].

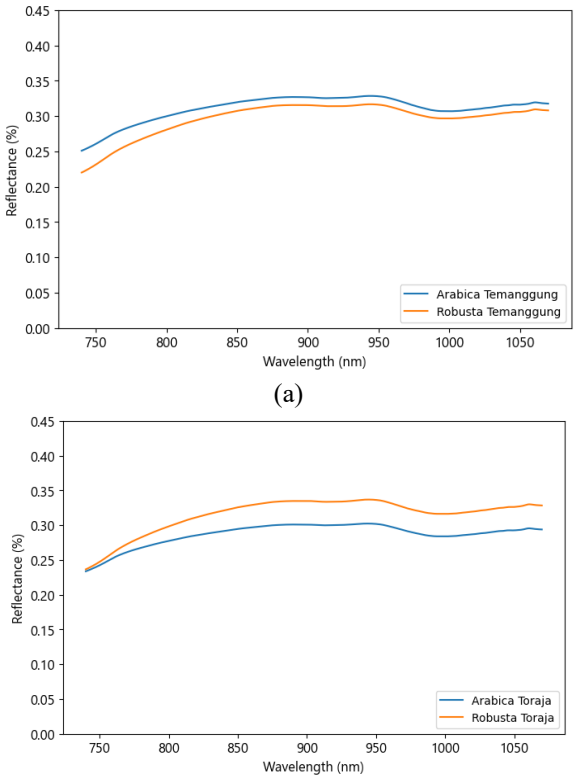

Fig 3. Original NIR Spectra of Arabica and Robusta coffee: (a) Temanggung and (b) Toraja

3.2 Results of PCA

PCA reduces variables of the spectroscopic data and generates new variables, called PCs, that are linear combinations of the original data and perform the maximum explained variance within the data [27]–[29]. The number of variables in the new data extracted by PCA may vary depending on the sorted eigenvalues. These eigenvalues are presented as percentages, also often called percent explained variance. We can also practically determine the number of PCs if the total population variance is over 80 to 90% [30]. The closer to 100 % of the total proportion of explained variance, the more faithful PCA represents the linear combinations of original data.

We determined the number of PCs used for further analyses was 3, which means three variables were obtained to perform LDA and ANN models. In Temanggung samples, the first three PCs accounted for 95.8% (PC1), 3.97% (PC2), and 0.2% (PC3) of the 99.97% explained total variance, while in Toraja samples, the PCA score accounted for 99.97% of the explained total variance from PC1: 96.88%, PC2: 2.95% and PC3: 0.14%. We also performed the separation between samples by plotting the two-dimensional PCA score plot (PC1 vs PC2), as depicted in Fig 4. Along PC1 and PC2, all samples seemed more dispersed and overlapped, but the distinguishment of Toraja samples showed satisfaction compared to Temanggung samples.

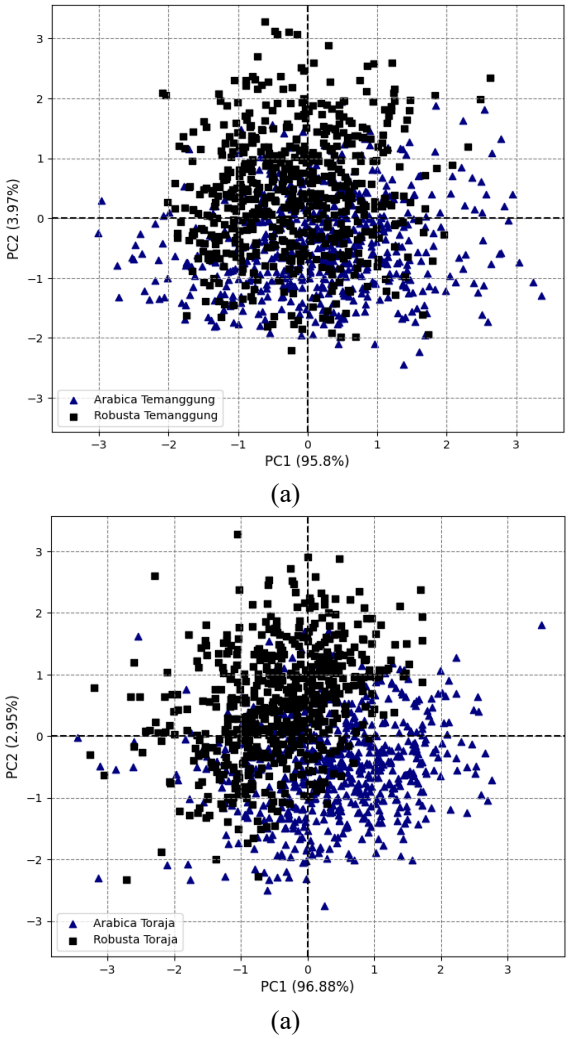

Fig 4. PCA score plots: (a) Temanggung and (b) Toraja

3.3 Classification Results of LDA

The new variables generated by PCA (the first three PCs) will be used as inputs to develop an LDA model. LDA provides more class separability and draws a decision boundary between the sample classes [31]. LDA computes an optimal data projection by minimizing within-class distance and maximizing between-class distance. Therefore, LDA is able to guarantee maximum class separation [32]–[34]. The projected data to visualize the discrimination between Arabica and Robusta coffee classes was performed in the LDA plot, see Fig 5. We still saw some Robusta samples were correctly predicted as Arabica and uncorrecly assigned as Robusta samples. Likewise, Arabica samples were predicted correctly and incorrectly.

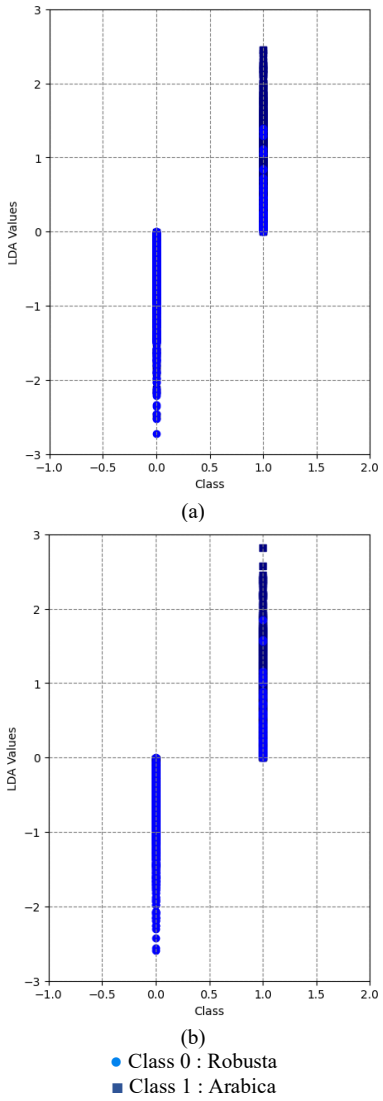

Fig 5. LDA plots: (a) Temanggung, and (b) Toraja

The confusion matrix (Table 2) evaluated the LDA model and figured out the classification results. The LDA model produced an accuracy value of >75% for Temanggung and >80% for Toraja. The classification error value generated for Temanggung reached below 25% and 20% for Toraja. The specificity, precision, recall, and F1-score matrices also showed a value of >75% for Temanggung and >80% for Toraja.

3.4 Classification Results of ANN

This ANN classifier was categorized as supervised learning in which certain sample groups were determined based on feature data and its label [35], [36]. Fig 6 depicts the curves representing how this model works in the cross-validation process and to see the loss during training. The loss quantifies the discrepancy between the predicted and actual classes, while the accuracy measures the agreement. The ideal loss curve decreases and tends to flatten as it approaches 0.00 on the y-axis. Both loss curves did not closely approach the threshold value of 0.00 until the epochs ended, indicating some errors were generated during the training of the samples. The loss curve for Toraja obtained a lower error or closer to the threshold value of 0.00 than for Temanggung. A good model must achieve an accuracy value close to 1.00 in the accuracy curve. The accuracy curve of the Toraja represented more promising. It increased closer to a threshold value. All loss curves decreased to a stability point with a moderate gap between the two final curves. Continuing the training or more than 100 epochs to train the model may lead to 'overfit'.

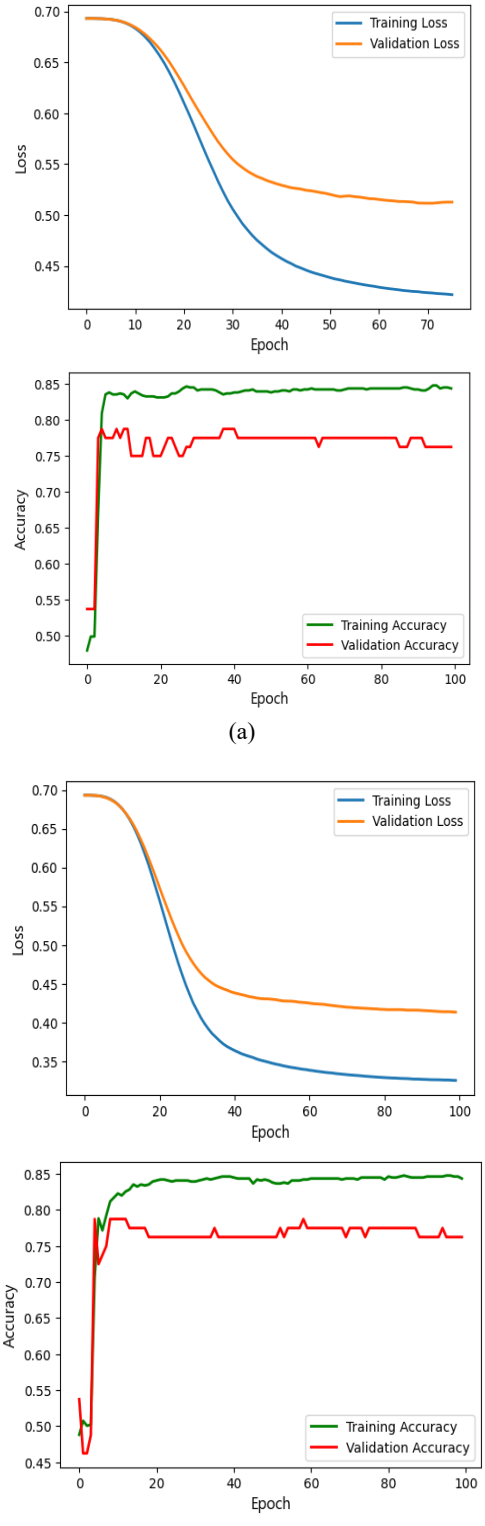

Fig 6. Loss and accuracy curves: (a) Temanggung and (b) Toraja

This study used three PCA-generated variables as inputs in the ANN model. The classification results were evaluated using a confusion matrix and performance metrics in Table 3. The results performed accuracy, specificity, precision, recall, F1-score, and error were not significantly different if we compared with the LDA results. Using a portable NIR spectrometer and ANN model, Toraja samples performed better in distinguishing coffee species compared to Temanggung samples.

The good ability of the ANN model to distinguish coffee species was also visualized in the ROC curve and AUC values (Fig 7). In Toraja samples, the ROC curves of Robusta and Arabica (Fig 7b) were closer to point (0,1) and consistently stayed away from above the ascending diagonal line. The AUC values showed closer to a value of 1, indicating that using this spectrometer and ANN model obtained better performance.

Table 2. Performance metrics of LDA

Samples			Confusion matrix		k-fold CV	AC*	SP*	PR*	RE*	FS*	E*
Toraja			R	A							
	Training	R	324	101	77.5±7.8	77.9	76.2	79.7	80.8	80.2	22.1
		A	77	302			79.7	76.2	74.9	75.6	
	Testing	R	164	47		79.3	77.7	81.1	82.4	81.7	20.7
		A	35	150			81.1	77.7	76.1	76.9	
Toraja			R	A							
	Training	R	336	67	83.3±6.1	83.3	83.4	83.2	83.4	83.3	16.7
		A	67	333			83.2	83.4	83.2	83.3	
	Testing	R	162	39		81.6	80.6	82.5	82.6	82.6	18.4
		A	34	161			82.5	80.6	80.5	80.5	

* All units in percentage

Table 3. Performance metrics of ANN

Samples			Confusion matrix		k-fold CV	AC*	SP*	PR*	RE*	FS*	E*
Toraja			R	A							
	Training	R	356	116	79.3 ± 5.6	80.4	70.9	75.4	89.9	82.0	19.6
		A	40	283			89.9	87.6	70.9	78.4	
	Testing	R	169	57		79.0	71.6	74.8	86.7	80.3	21.0
		A	26	144			86.7	84.7	71.6	77.6	
Toraja			R	A							
	Training	R	356	84	83.9 ± 3.1	83.6	78.9	80.9	88.1	84.4	16.4
		A	48	315			88.1	86.8	78.9	82.7	
	Testing	R	159	38		81.3	81.1	80.7	81.5	81.1	18.7
		A	36	163			81.5	81.9	81.1	81.5	

* All units in percentage

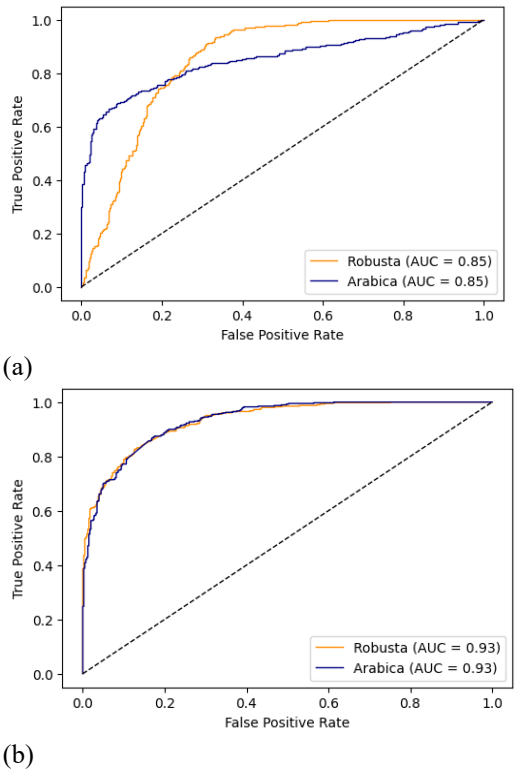

Fig 7. AUC-ROC curves of the ANN model: (a) Temanggung and (b) Toraja

4 Conclusions

The portable NIR Spectrometer associated with qualitative chemometric and artificial neural network models is moderately effective in distinguishing Arabica and Robusta coffee species. It is a promising technique that can be applied routinely for detecting and controlling coffee samples. It will reduce the time of analyses, be non-destructive without prior sample preparation, and provides in situ and real-time data acquisition. It can practically be applied at industrial levels to migrate from an analytical instrument to a reliable-lightweight tool. It may also become one of the promising methods for spectroscopy-based real-time food classification with a comparable result. Howefer, the accuracy of measurement is still one of the key challenges. Therefore, the implementation from different samples combined chemometric and machine learning models needs to be considered for further studies to guarantee consistency of generalization, not only to address classification problems but also to detect quality parameters of food.

References

[1] BPS, "Statistik Kopi Indonesia 2021," Jakarta, 2021.

[2] P. R. A. B. Toledo, L. Pezza, H. R. Pezza, and A. T. Toci, "Relationship Between the Different Aspects Related to Coffee Quality and Their Volatile Compounds," *Compr. Rev. Food Sci. Food Saf.*, vol. 15, no. 4, pp. 705–719, 2016, doi: 10.1111/1541-4337.12205.

[3] I. Hečimović, A. Belščak-Cvitanović, D. Horžić, and D. Komes, "Comparative study of polyphenols and caffeine in different coffee varieties affected by the degree of roasting," *Food Chem.*, vol. 129, no. 3, pp. 991–1000, 2011, doi: 10.1016/j.foodchem.2011.05.059.

[4] N. Liang, S. Sun, C. Zhang, Y. He, and Z. Qiu, "Advances in infrared spectroscopy combined with artificial neural network for the authentication and traceability of food," *Crit. Rev. Food Sci. Nutr.*, vol. 62, no. 11, pp. 2963–2984, 2022, doi: 10.1080/10408398.2020.1862045.

[5] M. R. Baqueta, A. Coqueiro, and V. Patricia, "Brazilian Coffee Blends: A Simple and Fast Method by Near-Infrared Spectroscopy for the Determination of the Sensory Attributes Elicited in Professional Coffee Cupping," *J. Food Sci.*, vol. 0, no. 0, 2019, doi: 10.1111/1750-3841.14617.

[6] H. T. M. Tran, C. A. C. Vargas, L. Slade Lee, A. Furtado, H. Smyth, and R. Henry, "Variation in bean morphology and biochemical composition measured in different genetic groups of arabica coffee (Coffea arabica L.)," *Tree Genet. Genomes*, vol. 13, no. 3, 2017, doi: 10.1007/s11295-017-1138-8.

[7] A. Khoshroo, A. Arefi, A. Masoumiasl, and G.-H. Jowkar, "Classification of Wheat Cultivars Using Image Processing and Artificial Neural Networks .," *Academia*, vol. 2, no. 1, pp. 17–22, 2014.

[8] A. Taner, Y. B. Oztekin, A. Tekgüler, H. Sauk, and H. Duran, "Classification of Varieties of Grain Species by Artificial Neural Networks," *Agronomy*, vol. 8, no. 7, 2018, doi: 10.3390/agronomy8070123.

[9] K. Kiliç, I. H. Boyaci, H. Köksel, and I. Küsmenoglu, "A classification system for beans using computer vision system and artificial neural networks," *J. Food Eng.*, vol. 78, no. 3, pp. 897–904, 2007, doi: 10.1016/j.jfoodeng.2005.11.030.

[10] M. Meenu, C. Kurade, B. C. Neelapu, S. Kalra, H. S. Ramaswamy, and Y. Yu, "A concise review on food quality assessment using digital image processing," *Trends Food Sci. Technol.*, vol. 118, no. PA, pp. 106–124, 2021, doi: 10.1016/j.tifs.2021.09.014.

[11] V. D. C. Martins *et al.*, "Fraud investigation in commercial coffee by chromatography," *Food Qual. Saf.*, vol. 2, no. 3, pp. 121–133, 2018, doi: 10.1093/fqsafe/fyy017.

[12] C. A. Teixeira Dos Santos, M. Lopo, R. N. M. J. Páscoa, and J. A. Lopes, "A review on the applications of portable near-infrared spectrometers in the agro-food industry," *Appl. Spectrosc.*, vol. 67, no. 11, pp. 1215–1233, 2013, doi: 10.1366/13-07228.

[13] I. Esteban-Díez, J. M. González-Sáiz, C. Sáenz-González, and C. Pizarro, "Coffee varietal differentiation based on near infrared spectroscopy," *Talanta*, vol. 71, no. 1, pp. 221–229, 2007, doi: 10.1016/j.talanta.2006.03.052.

[14] E. K. Kemsley, S. Ruault, and R. H. Wilson, "Discrimination between Coffea arabica and

Coffea canephora variant robusta beans using infrared spectroscopy," *Food Chem.*, vol. 54, no. 3, pp. 321–326, 1995, doi: https://doi.org/10.1016/0308-8146(95)00030-M.

[15] W. Zheng, X. Fu, and Y. Ying, "Spectroscopy-based food classification with extreme learning machine," *Chemom. Intell. Lab. Syst.*, vol. 139, pp. 42–47, 2014, doi: 10.1016/j.chemolab.2014.09.015.

[16] R. M. Correia *et al.*, "Portable near infrared spectroscopy applied to quality control of Brazilian coffee," *Talanta*, vol. 176, no. April 2017, pp. 59–68, 2018, doi: 10.1016/j.talanta.2017.08.009.

[17] M. R. Baqueta, A. Coqueiro, P. H. Março, and P. Valderrama, "Multivariate classification for the direct determination of cup profile in coffee blends via handheld near-infrared spectroscopy," *Talanta*, vol. 222, no. July 2020, p. 121526, 2021, doi: 10.1016/j.talanta.2020.121526.

[18] A. Géron, *Hands-on Machine Learning whith Scikit-Learing, Keras and Tensorfow*. 2019.

[19] I. Saputra and D. A. Kristiyanti, *Machine Learning untuk Pemula*. Bandung: Informatika, 2022.

[20] A. Kulkarni, D. Chong, and F. A. Batarseh, "5 - Foundations of data imbalance and solutions for a data democracy," F. A. Batarseh and R. B. T.-D. D. Yang, Eds. Academic Press, 2020, pp. 83–106.

[21] D. Ballabio and R. Todeschini, "Chapter 4 - Multivariate Classification for Qualitative Analysis," D.-W. B. T.-I. S. for F. Q. A. and C. Sun, Ed. San Diego: Academic Press, 2009, pp. 83–104.

[22] M. T. Ahad, Y. Li, B. Song, and T. Bhuiyan, "Comparison of CNN-based deep learning architectures for rice diseases classification," *Artif. Intell. Agric.*, vol. 9, pp. 22–35, 2023, doi: 10.1016/j.aiia.2023.07.001.

[23] F. Gordillo-Delgado, A. Bedoya, and E. Marín, "Study of the Pigments in Colombian Powdered Coffee Using Photoacoustic Spectroscopy," *Int. J. Thermophys.*, vol. 38, no. 1, pp. 1–8, 2017, doi: 10.1007/s10765-016-2144-z.

[24] N. Farhaty and Muchtaridi, "Tinjauan Kimia dan Aspek Farmakologi Senyawa Asam Klorogenat pada Biji Kopi: Review," *Farmaka Suplemen*, vol. 14, no. 1, pp. 214–227, 2019.

[25] J. S. Ribeiro, M. M. C. Ferreira, and T. J. G. Salva, "Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy," *Talanta*, vol. 83, no. 5, pp. 1352–1358, 2011, doi: 10.1016/j.talanta.2010.11.001.

[26] S. S. N. Chakravartula, R. Moscetti, G. Bedini, M. Nardella, and R. Massantini, "Use of convolutional neural network (CNN) combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee," *Food Control*, vol. 135, no. January, p. 108816, 2022, doi: 10.1016/j.foodcont.2022.108816.

[27] H. Zhao, B. Guo, Y. Wei, and B. Zhang, "Near infrared reflectance spectroscopy for determination of the geographical origin of wheat," *Food Chem.*, vol. 138, no. 2–3, pp. 1902–1907, 2013, doi: 10.1016/j.foodchem.2012.11.037.

[28] Y. He, S. Feng, X. Deng, and X. Li, "Study on lossless discrimination of varieties of yogurt using the Visible/NIR-spectroscopy," *Food Res. Int.*, vol. 39, no. 6, pp. 645–650, Jul. 2006, doi: 10.1016/j.foodres.2005.12.008.

[29] J. N. Miller and J. C. Miller, *Statistics and Chemometrics for Analytical Chemistry*. Essex, England: Pearson Education Limited, 2010.

[30] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis, 6th edition*. Essex, England: Pearson Education Limited, 2014.

[31] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis-a brief tutorial," *Inst. Signal Inf. Process.*, vol. 18, pp. 1–8, 1998.

[32] J. Ye and S. Ji, "Discriminant Analysis for Dimensionality Reduction: An Overview of Recent Developments," *Biometrics Theory, Methods, Appl.*, pp. 1–19, 2009, doi: 10.1002/9780470522356.ch1.

[33] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AI Commun.*, vol. 30, no. 2, pp. 169–190, 2017, doi: 10.3233/AIC-170729.

[34] C. H. Park and H. Park, "A comparison of generalized linear discriminant analysis algorithms," *Pattern Recognit.*, vol. 41, no. 3, pp. 1083–1097, 2008, doi: 10.1016/j.patcog.2007.07.022.

[35] R. Vadivambal and D. S. Jayas, *Bio-Imaging: Principles, Techniques, and Applications*. Florida: Taylor & Francis Group, LLC, 2016.

[36] H. O. Velesaca, P. L. Suárez, R. Mira, and A. D. Sappa, "Computer vision based food grain classification: A comprehensive survey," *Comput. Electron. Agric.*, vol. 187, no. September 2020, 2021, doi: 10.1016/j.compag.2021.106287.