International Journal of Agriculture and Biosciences 2025 14(6): 1291-1301.
https://doi.org/10.47278/journal.ijab/2025.110

RESEARCH ARTICLE
eISSN: 2306-3599; pISSN: 2305-6622

Authentication of Robusta Coffee Origins by Shortwave NIR Spectroscopy Coupled with Dimensionality Reduction and Neural Networks

Agus Dharmawan [1] and Rudiati Evi Masithoh [2,*]

[1]Department of Agricultural Engineering, Faculty of Agricultural Technology, Jember University, Indonesia
[2]Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Yogyakarta, Indonesia 55281
*Corresponding author: evi@ugm.ac.id

ABSTRACT

Non-destructive techniques such as spectroscopy are widely used to authenticate the geographical origins of food and agricultural products. This study presents an integrated approach using shortwave near-infrared (SWNIR) spectroscopy, dimensionality reduction, and artificial neural networks (ANN) to authenticate Robusta coffee beans from four regions in Indonesia: Temanggung, Toraja, Dampit, and Lampung. Spectral data collected in the 954–1700 nm range were transformed using three linear dimensionality reduction methods—principal component analysis (PCA), partial least squares (PLS), and linear discriminant analysis (LDA). The resulting feature sets were used to train ANN classifiers. PCA, PLS, and LDA score plots demonstrated clear clustering among coffee origins. Results show that the LDA–ANN combination achieved the highest classification accuracy of 100%, along with perfect values for precision, recall, specificity, and F1-score. In contrast, PCA–ANN and PLS–ANN reached accuracies of 97.9 and 96.2%, respectively. The ROC and AUC analysis further confirmed the superior separability of LDA-based classification, showing no overlap between sample classes. These findings highlight the potential of SWNIR spectroscopy combined with LDA and ANN for rapid, reliable, and non-destructive geographical authentication of Robusta coffee.

Keywords: Artificial neural network; Geographical origin; PCA; PLS; LDA

Article History
Article # 25-219
Received: 28-Apr-25
Revised: 07-Jun-25
Accepted: 05-Jul-25
Online First: 18-July-25

INTRODUCTION

Coffee is one of the most widely consumed beverages worldwide and plays a vital role in the global economy (Torga & Spers, 2020). Among the two most commercially cultivated species, Arabica (Coffea arabica) and Robusta (Coffea canephora); Robusta has higher caffeine content and stronger taste (Davila & Sîrbu, 2021). However, the quality of coffee varies according to several factors, such as post-harvest processing methods (Velásquez & Banchón, 2023), genetic variation, and environmental conditions (Ahmed et al., 2021). These factors influence coffee beans' physical and chemical composition, determining sensory attributes like aroma, flavor and appearance. Consequently, coffee originating from different geographic regions often commands different market values and consumer preferences due to its distinct characteristics (Abdu & Mutuku, 2021). Consumer awareness regarding product authenticity has increased significantly in recent years (Chousou & Mattas, 2019). Verification of the geographical origin of beans is crucial for the coffee industry in maintaining regional branding (Trihartono, 2022) and reducing the risk of mislabeling, which can impact consumer trust (Nunes et al., 2021). Ensuring the authenticity of coffee origin promotes fair trade, allows producers to charge premium prices (Wahyudi et al., 2020) and supports sustainable farming (Pratama & Wisika, 2022). Therefore, it is important to find efficient and reliable tools to authenticate the origins of coffee.

Traditional methods for determining coffee origin can be done using sensory evaluation or conventional chemical analysis (Bessadaet al., 2018; Poláková et al., 2023). However, the methods are often time-consuming, destructive and require high equipment investment and analytical costs. In contrast, spectroscopic techniques offer a non-destructive, rapid and cost-effective alternative for

Cite this Article as: Dharmawan A and Masithoh RE, 2025. Authentication of robusta coffee origins by shortwave NIR spectroscopy coupled with dimensionality reduction and neural networks. International Journal of Agriculture and Biosciences 14(6): 1291-1301. https://doi.org/10.47278/journal.ijab/2025.110

A Publication of Unique Scientific Publishers

analyzing the quality parameters of food and agricultural products (Nawrocka & Lamorska, 2016). One of the spectroscopic methods is Near-Infrared (NIR) spectroscopy, which has been proven as a powerful analytical tool for qualitative and quantitative assessment of coffee (Munyendoet al., 2022). It captures the vibrational overtones and combinations of molecular bonds—particularly O–H, C–H, and N–H— which reflect chemical constituents such as moisture, caffeine, chlorogenic acid, carbohydrate, lipids and sugars (Caporaso et al., 2018; Wei et al., 2021; Ayu et al., 2024). These spectral features serve as chemical fingerprints that distinguish coffee origins and cultivars (Malta et al., 2020).

Several studies have reported on the use of NIR spectroscopy and various chemometric techniques for coffee classification. Giraudo et al. (2019) employed Fourier Transform NIR (FTNIR) spectroscopy combined with partial least squares discriminant analysis (PLS-DA) to classify green coffee beans based on continents and countries bases. Kurniawan et al. (2019) used FTNIR to discriminate Arabica Java coffee using principal component analysis (PCA) and discriminant analysis. Similarly, Yusmanizar & Munawar (2021) classified Arabica and Robusta coffee based on NIR spectra using linear discriminant analysis (LDA) and support vector machines (SVM). Other studies by Guerrero-Peña et al. (2023) employed PCA, hierarchical clustering (HC), and Soft Independent Modelling of Class Analogy (SIMCA) for the classification of regions and varieties. Pahlawan & Masithoh (2022) employed Visible-NIR spectroscopy and PLS-DA for classification of Arabica and Robusta roasted coffee. These studies, nonetheless, rely on chemometric approaches to classify samples based on spectral features that may not fully reveal the nonlinear relationships embedded in complex spectral data (Yang et al., 2019). On the other hand, one of the nonlinear models, namely artificial neural networks (ANN), is capable of solving both supervised and unsupervised classification problems (Bhagya Raj & Dash, 2022). ANN models have been applied to food quality evaluation, including classification and prediction (Ikram et al., 2024), as well as food traceability (Liang et al., 2022).

Despite the success of such methods for food products, limited studies have explored the combination of ANN models and multiple dimensionality reduction techniques, such as PCA, PLS, and LDA, for classifying Robusta coffee based on shortwave NIR (SWNIR) spectral data. Dimensionality reduction is a crucial pre-processing step for analyzing spectroscopic data, as it reduces the number of variables while retaining the most informative features (Nanga et al., 2021). In short, PCA is best for general dimensionality reduction without considering class labels; PLS is used for predictive modeling when class information is crucial, while LDA surpasses classification by maximizing class separability (Ayesha et al., 2020). Previous studies have reported on the implementation of dimensionality reduction and classification techniques. For instance, PCA with various classifiers such as SVM, k-NN, and BPNN were used to distinguish between Arabica and Robusta coffee using FTIR spectra (Zheng et al., 2014). FTNIR combined with multiple classifiers such as Self-Organizing Maps (SOM), SIMCA, and PLS-DA were used to discriminate Robusta coffee cultivar (Luna et al., 2017). Dharmawan et al. (2023b) used PCA with a multilayer perceptron ANN (MLP-ANN) to identify Arabica coffee origins. Nevertheless, studies comparing the effectiveness of various dimensionality reduction methods combined with ANN models for Robusta coffee origin classification are limited. Current research has focused on Arabica and green coffee beans, while Robusta coffee from diverse Indonesian origins has not been sufficiently studied. Moreover, the previous studies employed spectroscopic data obtained from NIR instruments at wavelengths 1000-2500 nm (Sim et al., 2024) or IR instruments at wavelengths 2500-4000 nm (Obeidat et al., 2018). The instrument used in this study is fiber optic spectroscopy which is portable and low-cost, which is affordable for small-scale coffee farmers or industries for their quality evaluation. The instrument has been used for grain such as coffee, soybean, and cocoa bean (Priambodo et al., 2022; Dharmawan et al., 2023; Abadi et al., 2024). Therefore, this study aims to evaluate the classification performance of ANN models using SWNIR spectra data 1000–1700 nm, which was reduced by PCA, PLS, and LDA dimensionality reduction techniques. Robusta coffee samples from four locations in Indonesian, i.e., Temanggung, Toraja, Dampit, and Lampung, were analyzed to identify which technique combination best provides the most accurate and reliable authentication model to differentiate the coffee samples by geographic origin.

MATERIALS & METHODS

Spectral Acquisition

A total of 2,400 spectral readings were collected from 600 Robusta coffee beans (150 beans per origin), representing four geographic origins in Indonesia: Temanggung, Toraja, Dampit, and Lampung (Fig. 1). All beans were obtained from dry-processed coffee and were manually selected to ensure uniformity, with damaged beans removed. Prior to spectral acquisition, the beans were cleaned of residual endocarp (parchment layer) and surface contaminants.

Spectral data were acquired using a shortwave near-infrared (SWNIR) spectrometer (Ocean Optics, Orlando, FL, USA) with a wavelength range of 954–1700 nm at 6 nm intervals, resulting in 125 spectral variables after deleting the initial and last spectra, which were noisy. The setup included a tungsten halogen light source (HL-2000-HP-FHSA, 360–2400 nm) and a reflectance fiber optic probe (QR400-7-VIS–NIR, Ocean Optics), Fiber Connector SMA 905, nominal bulb power 20 W, typical output power 8.4 mW), and fiber optic cable reflection probe (Type: QR400-7-VIS-NIR Ocean Optics, wavelength range: 400-2100nm) (Prasetyo et al., 2024). Fig. 2 illustrates the spectral acquisition setup used in this study, including probe alignment and sample positioning. White reference spectra were recorded using a white ceramic standard to ensure calibration, while black reference spectrum was obtained by switching off the light source (Dharmawan et al., 2023a). Raw spectral data were saved in .csv format for subsequent analysis.

Dimensional Reduction

Three linear dimensional reduction techniques—Principal Component Analysis (PCA), Partial Least Squares (PLS), and linear discriminant analysis (LDA)—were

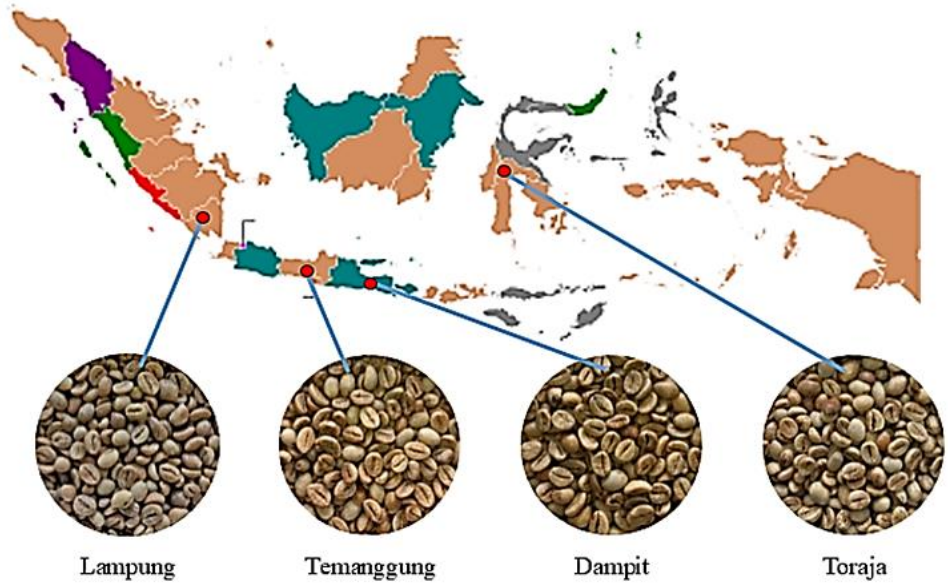


Fig. 1: Locations of origin of four Robusta coffee beans with their corresponding coffee images.

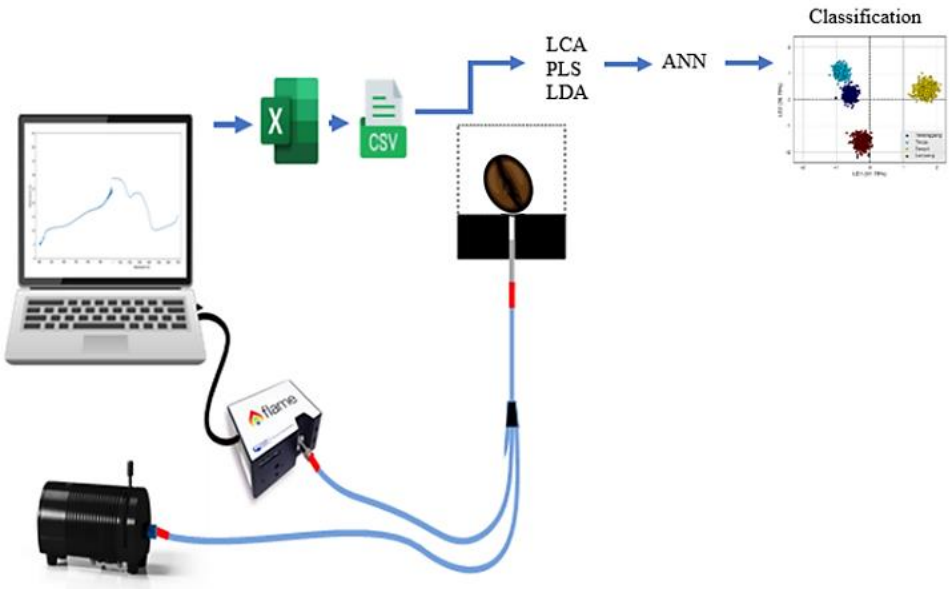


Fig. 2: Schematic of coffee reflectance acquisition using SWNIR spectroscopy.

employed to reduce the high-dimensional spectral data into low-dimensional feature sets. PCA is an unsupervised method, while PLS and LDA are supervised methods that utilize a response Y-matrix representing the four Robusta coffee origins (Temanggung, Toraja, Dampit and Lampung). Each method transformed the spectral data into new variables (scores), which are linear combinations of the original spectra (Ayesha et al., 2020). These scores were then scaled using Z-score normalization. The dimensionality-reduced data were evaluated through score plot visualizations and the explained variance proportions. PCA and PLS retained the first four components, accounting for more than 95% of the total variance. Due to class limitations, LDA yielded a three-dimensional score matrix (number of classes—1).

Development of ANN Classifier

A Multilayer Perceptron (MLP) architecture was developed for classification, consisting of an input layer, two hidden layers, and an output layer. The number of input neurons varied depending on the dimensional reduction method: PCA and PLS used four nodes; LDA used three nodes) as shown in Table 1. Both hidden layers used the Rectified Linear Unit (ReLU) activation function with Glorot weight initialization. In contrast, the output layer used Softmax activation and He initialization to classify the samples into one of the four Robusta coffee origins. Target labels were encoded using one-hot encoding, assigning binary values (0 or 1) to each class: 0 = Temanggung, 1 = Toraja, 2 = Dampit, 3 = Lampung.

K-fold cross-validation was applied to evaluate model generalizability. The dataset was divided into training (2/3) and testing (1/3) sets. During each k iteration, one subset was used for validation, while the remaining k–1 subset was used for training. The model's performance was assessed by calculating the mean and standard deviation of accuracy across all iterations (Xiong et al., 2020).

Classifier Performance

Loss and Accuracy Curves

To evaluate the model's learning behavior and generalization ability, accuracy and loss curves were plotted using the training data obtained through cross-validation (Novtahaning et al., 2022). The accuracy curve reflects the

Table 1: Structure of ANN classifier

Name of layers	Number of Neurons (nodes)			Activation function	Weight initialization	Other parameters
	PCA	PLS	LDA			
Input layer	4	4	3			• Loss function = Categorical cross-entropy
Hidden layer 1	6	6	6	ReLU*	Glorot	• Optimizer = Adam
Hidden layer 2	6	6	6	ReLU	Glorot	• Validation control = Metric 'Accuracy'
Output layer	4	4	4	Softmax	He	• Number of epochs = 100
						• Batch size = 100
						• Callback function = EarlyStopping

*ReLU = rectified linear unit

model's ability to correctly classify samples by comparing the predicted outputs with actual labels. In contrast, the loss curve quantifies the classification error during training. A high loss value indicates poor model performance due to significant misclassification, while high accuracy corresponds to a lower error rate and better predictive performance. These visualizations provide insight into whether the model is learning effectively or overfitting the training data.

Confusion Matrix and Classification Metrics

The confusion matrix in Fig. 3 was used to evaluate the performance of the classification models. A confusion matrix provides a summary of a classifier's predictive performance by comparing the predicted class labels with the actual (true) labels of the dataset (Ahad et al., 2023). In this matrix, the predicted targets represent the class labels assigned by the model, while the actual targets represent the original class labels of the coffee samples.

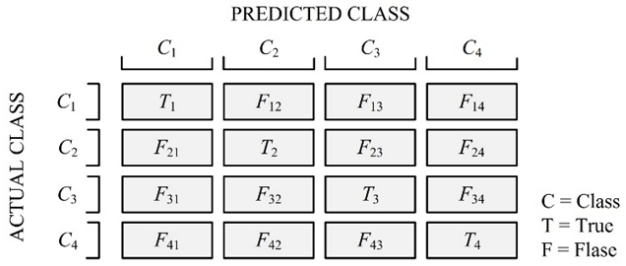


Fig. 3: Four-class confusion matrix.

These values are extracted from Table 2 to assess the classification accuracy and error types. The confusion matrix serves as a reliable diagnostic tool for understanding model performance beyond overall accuracy, especially in multi-class classification tasks (Géron, 2019).

The performance of the classification model was evaluated using several standard metrics based on the confusion matrix (De Diego et al., 2022). Accuracy (AC), or the classification rate or non-error rate, represents the proportion of correctly classified samples out of the total number of samples calculated as in Equation 1. The error rate (ER) is the complementary accuracy index, indicating the proportion of misclassified samples, calculated as shown in Equation 2. An ideal classifier yields an accuracy close to 100% and an error rate near 0%.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{Equation 1}$$

$$Error\_rate = \frac{FP+FN}{TP+TN+FP+FN} \quad \text{Equation 2}$$

The Recall (RE), also known as sensitivity, measures the model's ability to correctly identify positive samples. Precision (PR) represents the proportion of correctly identified positive samples out of all samples predicted as positive. Specificity (SP) indicates the model's ability to correctly classify negative samples. The RE, PR, and SP are calculated using Equation 3, 4, and 5.

$$Recall = \frac{TP}{TP+FN} \quad \text{Equation 3}$$

$$Precision = \frac{TP}{TP+FP} \quad \text{Equation 4}$$

$$Specifity = \frac{TN}{TP+FP} \quad \text{Equation 5}$$

An additional metric used in this study is the F1-score, which is the harmonic mean of Precision and Recall. It provides a balanced classification performance measure, especially in imbalanced datasets.

$$F1 - score = 2x\frac{PrecisionxRecall}{Precision+Recall} \quad \text{Equation 5}$$

In general, Recall, Precision and Specificity are considered good when they approach 100%. For the F1-score, a perfect classification corresponds to a value of 1.0, while a score closer to 0.0 indicates poor performance.

ROC and AUC Value

The classifier's performance was further assessed by visualizing the Receiver Operating Characteristic (ROC) curve and calculating the Area Under the Curve (AUC) (Aguilar-Ruiz & Michalak, 2022). The ROC curve is a graphical representation that plots the True Positive Rate (TPR) (or Sensitivity/Recall) on the y-axis against the False Positive Rate (FPR) on the x-axis (Carrington et al., 2022). The True Negative Rate (TNR), also referred to as specificity, represents the proportion of negative samples correctly classified as negative. The False Positive Rate (FPR) is the proportion of negative samples incorrectly classified as positive, and it is mathematically expressed as [1 - TNR]. Consequently, the ROC curve visualizes Sensitivity (Recall) versus [1 - specificity].

A classification model is considered poor if its ROC curve is close to the baseline, which is represented by the diagonal line that connects the points (0, 0) and (1, 1). A model is superior if the curve is near the top-left corner, approaching the point (0,1) (Bowers & Zhou, 2019).

The AUC quantifies the classifier's overall performance by summarizing the ROC curve into a single numerical value. The AUC ranges from 0 to 1, where a higher value indicates better model performance. An AUC value of 0.5 suggests a model with no discrimination ability (similar to random guessing), while an AUC value approaching 1.0 indicates a strong ability to separate classes (Bowers & Zhou, 2019).

Table 2: Determination of TP, TN, FP, and FN

Class	C1	C2	C3	C4
	Temanggung	Toraja	Dampit	Lampung
TP	T1	T2	T3	T4
TN	T2 + T3 + T4 + F23 + F24 + F32 + F34 + F42 + F43	T1 + T3 + T4 + F13 + F14 + F31 + F41 + F34 + F43	T1 + T2 + T4 + F12 + F14 + F21 + F24 + F41 + F42	T1 + T2 + T3 + F12 + F13 + F21 + F23 + F31 + F32
FP	F21 + F31 + F41	F12 + F32 + F42	F13 + F23 + F43	F14 + F24 + F34
FN	F12 + F13 + F14	F21 + F23 + F24	F31 + F32 + F34	F41 + F42 + F43

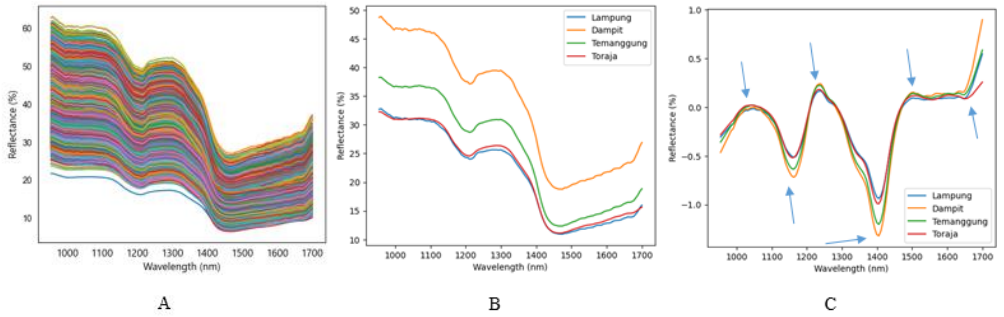


Fig. 4: Profile of Shortwave NIR of (a) all original spectra, (b) average spectra, and c) average Savitzky-Golay 1st Derivative PCA, PLS, and LDA.

Therefore, a good classification model will have an AUC greater than 0.5 and will lie above the diagonal baseline of the ROC curve (Aguilar-Ruiz & Michalak, 2022).

RESULTS & DISCUSSION

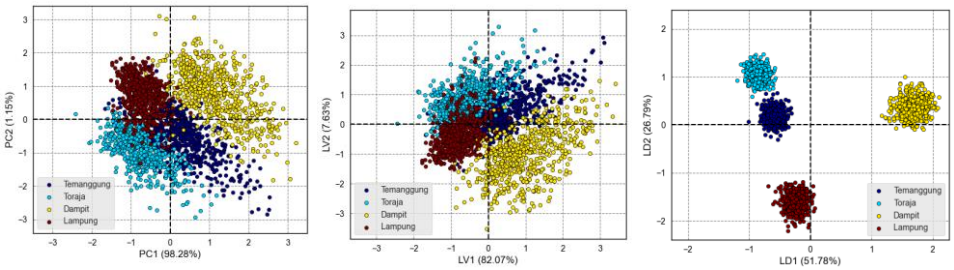
Spectra Visualization

Reflectance data for Robusta coffee from four regions in Indonesia were collected using a SWNIR spectrometer operating within the wavelength range of 954–1700nm, resulting in 126 spectral variables. Fig. 4a shows all original spectra, while Fig. 4b shows average spectra of four Robusta coffees for four regions. When analyzing the average spectra (Fig. 4b), the spectra were very similar over most of the range, except in some wavelength regions where the differences were noticeable. While the average spectra of the Robusta Lampung and Toraja overlap, Temanggung and Dampit show some differences in reflectance intensity. The difference between Lampung and Toraja varieties was not evident because their spectra overlapped, indicating that they had similar spectrochemical characteristics. Temanggung and Dampit coffee exhibit higher reflectance compared to Lampung and Toraja coffee across the wavelength range. This variation suggests differences in chemical composition among samples (Pandiselvam et al., 2022).

By applying Savitzky-Golay 1st Derivative to the original spectra several significant reflectance bands were more observable in the spectra between 900-1700nm. According to (Barbin et al., 2014), absorption of groups CH, C = C-H, H2O, CH3, CH2, RNH, and ROH related to carbohydrates, water, lipids, proteins, aromatics, chlorogenic acid, and caffeine present in those regions. Notably, there are prominent absorption bands corresponding to the CH stretching overtones at around 1200nm related to acidity (Ribeiro et al., 2011) and NH stretch of primary amines at 1400nm (Weyer & Lo, 2006). Lampung coffee exhibits the lowest reflectance or the highest absorbance compared to other varieties, particularly at 1450 nm, which is associated with the presence of carbohydrates, chlorogenic acids, and lipids in coffee (Baqueta et al., 2023) and OH stretching of water content (Barbin et al., 2014). Furthermore, the first overtone of O–H and N–H stretching, which relates to amino acids and chlorogenic acids, was observed at 1579nm (Chakravartula et al., 2022) and 1672 nm related to caffeine (Barbin et al., 2014). The spectral differences can be used to characterize and discriminate between coffee origins based on their unique NIR spectral fingerprints, which are directly linked to the functional groups involved (Giraudo et al., 2019). The bitterness of Robusta coffee, largely related to the levels of caffeine and chlorogenic acid (Ban et al., 2025), can be traced to these spectral features. This spectral information is valuable for understanding the chemical composition of Robusta coffee and can be applied in quality control.

To investigate the differences between Robusta coffee origins, PCA, PLS and LDA analyses were conducted. PCA was evaluated by plotting the first two principal components based on spectral characteristics. Fig. 5a displays a two-dimensional PCA score plot for four Robusta coffee origins (Temanggung, Toraja, Dampit, and Lampung), illustrating clear groupings based on their chemical and compositional similarities. The first two principal components, PC1 and PC2, explained 98.28% and 1.15% of the total variance, respectively, accounting for a total of 99.43% of the variance. Similar results reported by (Yusmanizar & Munawar, 2021) which use PC1 and PC2 to classify another variety of Indonesia coffee from Sumatra region. Fig. 5b shows the PLS score plot (Latent Variable 1 (LV1) vs. LV2). The separation of the coffee origins is evident, with the PLS components creating distinct boundaries between the groups, suggesting effective discrimination. LV1 accounted for 90.23% of the variance, while LV2 contributed 8.38%, resulting in an overall explained variance of 98.61%. The use of PLS to classify Indonesian coffee is more distinct that the one used to classify coffee from various countries such as Columbia, Ethiopia, India, Indonesia, and Nicaragua (Núñez et al., 2020).

Both PCA and PLS models successfully classify Indonesia coffee origin. The samples formed two distinct groups along PC1 and LV1: one group, composed of Temanggung and Dampit, had positive scores, while the other, consisting of Toraja and Lampung, had negative scores. This separation is evident in the two-dimensional plots, where at PC2 and LV2, the Temanggung and Toraja



A B C

Fig. 5: Plots of PC scores from original spectra based on (a) PCA, (b) PLS, and (c) LDA models.

samples shift to a negative score. At the same time, Dampit and Lampung move to a positive score. The spatial distribution of the samples suggests significant chemical compositional differences between these origins (Yusibani et al., 2023; Yani & Utomo, 2025), with the spectral characteristics of each origin contributing to the separation.

LDA demonstrated superior performance compared to PCA and PLS by providing the maximum separability of the classes and effectively drawing decision boundaries between the sample groups. Similar work was conducted by (Suhandy & Yulia, 2018) to classify Luwak coffee, obtaining 100% accuracy using LDA. LDA works by minimizing within-class variance and maximizing between-class variance (Ansari et al., 2025). As shown in Fig. 4c, the LDA score plot (LD1 vs. LD2) reveals a more clustered grouping of the coffee origins, with Dampit distinctly positioned away from the other samples along the positive score of LD1. The different origins—Temanggung, Toraja, and Lampung—are placed along the negative score of LD1, with further separation along LD2. Lampung, for example, is positioned away from the other samples along with the negative score of LD2, while Temanggung and Toraja are placed along with the positive score. The relatively small distance between Temanggung and Toraja suggests spectral similarities between these two origins. Overall, LDA provided a clearer separation and a more coherent interpretation of the data than PCA and PLS.

In addition to the score plots, PCA and PLS loading plots were examined to assess the contributions of different wavelengths to the variance in the data (Rismiwandira et al., 2020). The loading matrix helps to identify which variables contribute most to each component. Fig. 6 shows the loading plots for the first two PCs and LVs. Significant contributions are marked by values far from the x = 0 axis (either positive or negative), indicating that certain wavelengths strongly influence the principal components and, consequently, the separation between the samples. For PC1 (Fig. 6a), the significant wavelengths for sample distinction ranged from 940 to 1350nm (positive contribution) and from 1350 to 1700 nm (negative contribution). These contributions facilitated the separation of the Temanggung and Dampit samples from the Toraja and Lampung samples. The sample separation based on PC2 further refined this classification. Similarly, PLS LV1 (Fig. 6b) displayed high positive contributions between 940 and 1370nm, with negative contributions from the remaining wavelengths (1370–1700 nm), helping separate the Temanggung and Dampit samples from the Toraja and Lampung samples. The contribution of LV2 was particularly useful for differentiating the Temanggung and Dampit samples.

The PCA and PLS loading plots (Fig. 6) reveal distinctive peaks and valleys at specific wavelengths, indicating significant chemical differences between the coffee origins. At Fig. 6a, PC1 explained 98.28% of the total variance in the original spectra, with peaks around 1300, 1460 and 1650nm corresponding to chlorogenic acid (Correia et al., 2018), carbohydrates, chlorogenic acids, and lipids (Baqueta et al., 2023), water content (Barbin et al., 2014), and caffeine (Barbin et al., 2014). PC2 explained 1.15% of the variance, with additional peaks and valleys at 1100 and 1200nm representing the second overtone of the CH functional group (Correia et al., 2018). These wavelengths contribute to the spectral differences observed across the coffee samples.

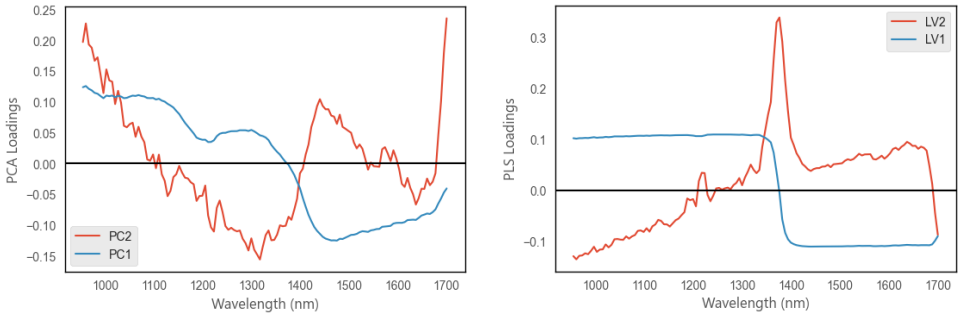
The LV1 from the PLS model (Fig. 6b) shows a flattened curve with an S-shape in the loadings, while LV2 exhibited high peaks at 1400 and 1650 nm. The reflectance at these wavelengths primarily arises from CH and OH vibrations in the chemical groups present in the coffee samples, providing key insights into the chemical composition of the Robusta coffee samples (Zhang et al., 2025).

The number of data variables in PCA and LDA was determined by evaluating the sorted eigenvalues. These eigenvalues are often expressed as percentages of the explained variance (Table 3). A higher cumulative percentage indicates a better model, as the eigenvalues closer to zero or with very low variance may suggest unimportant variables or potential linear dependencies in the data. These low-variance components likely do not contribute meaningfully to the data model, possibly reflecting extraneous factors influencing the observations (Brereton, 2009).

Table 3: The proportion of explained variance (%) to determine the number of reduced features

		Number of components (or k variables)			
		1	2	3	4
PCA	Explained variance	98.28	1.15	0.26	0.18
	Cummulative explained variance	98.28	99.43	99.69	99.87
PLS	Explained variance	90.23	8.38	0.85	0.22
	Cummulative explained variance	90.23	98.61	99.46	99.68
LDA	Explained variance	51.78	26.79	21.44	×
	Cummulative explained variance	51.78	78.57	100	×

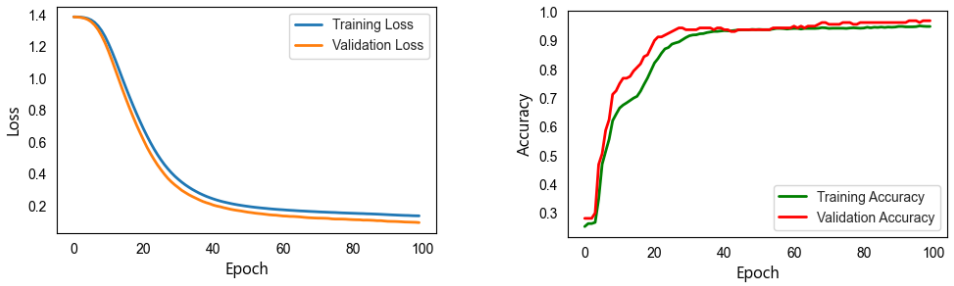
In contrast to PCA, the successive values of k in PLS do not necessarily decrease with each component calculated. This is because PLS models both the X (predictor) and Y (response) data simultaneously, using regression to establish a relationship between them (Brereton, 2009). To optimize the number of components (k) for classification studies, cross-validation is a commonly used method.



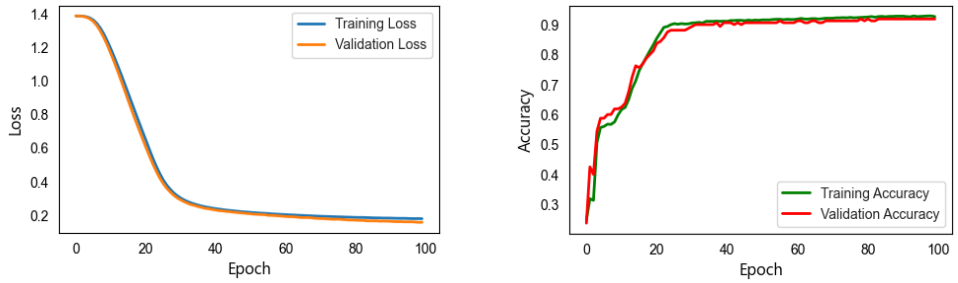
A B

Fig. 6: The loading plots based on (a) PCA and (b) PLS models.

PCA k-fold CV accuracy = 94.78% (±1.72%)



PLS k-fold CV accuracy = 92.10% (± 2.15%)



LDA k-fold CV accuracy = 97.58% (± 7.27%)

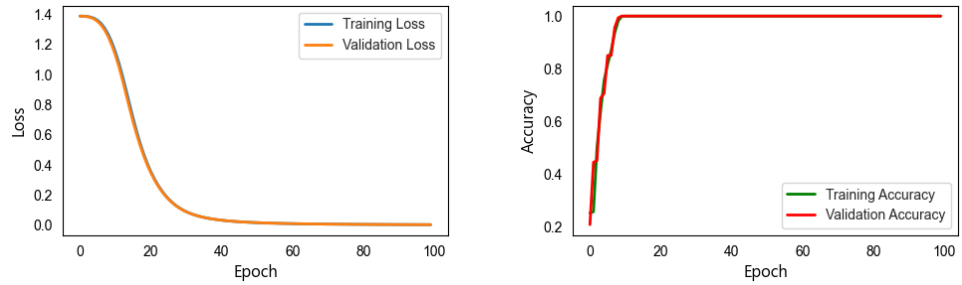


Fig. 7: Accuracy and loss curves based on PCA, PLS, and LDA models.

The correct number of PLS components is determined by selecting the k that results in the maximum accuracy, which is identified when the accuracy value stabilizes or increases with the addition of more components.

Integration of Dimensional Reduction and Artificial Neural Networks (ANN)

This section discusses the effectiveness of combining dimensional reduction techniques (PCA, PLS, and LDA) with artificial neural networks (ANN) in authenticating the origins of Robusta coffee. Fig. 7 illustrates the loss and accuracy curves during the cross-validation process, with the x-axis representing the number of epochs and the y-axis showing accuracy and loss percentages. These curves provide insights into the loss during training, quantifying the discrepancy between predicted and actual classes in the training set (Farhadpour et al., 2024). Accuracy, on the other hand, assesses the agreement between predicted and actual class. A lower loss signifies that the curve decreases significantly and flattens as it approaches 0.00 on the y-axis, while higher accuracy moves closer to 1.00.

Fig. 7 shows that the loss and accuracy curves for PCA, PLS, and LDA follow similar trends. However, there are distinguishing features based on the smoothness of the curves, the number of epochs required to reach stability, and the final values of the loss curves. PCA and PLS exhibited fewer smooth curves compared to LDA. The LDA dataset yielded superior sample authentication. The findings contradicted those of Sornam & Vanitha (2018), which reported better accuracy for PCA-based networks

compared to LDA-based networks. Furthermore, the loss curves for the validation dataset were consistently lower than those for the training dataset, indicating that the model successfully learned from the training set (Wang et al., 2022). The validation dataset, therefore, contributed to generalization. All three curves were categorized as "good fit," with training and validation losses decreasing and stabilizing, with a minimal gap between the two final loss values. Continuing training beyond this point could lead to overfitting, where the model performs well on training data but struggles with new data, as indicated by an increasing gap between the two curves. According to the loss curves in Fig. 7, no overfitting was observed, as the training and validation datasets were appropriately split.

A superior model will flatten the loss curve, ideally achieving a value of 0.00 on the y-axis. Compared to PCA and PLS, LDA performed excellently when combined with the ANN model. The loss curve for LDA reached the threshold value of 0.00, while the PLS and PCA curves did not closely approach this threshold until the final epochs. Additionally, the LDA accuracy curve reached a threshold of 1.00 more rapidly, stabilizing after approximately 10 epochs, demonstrating superior classification performance compared to PCA and PLS.

The classifier's performance was further evaluated using several metrics. These metrics help assess how well the classifier performs in terms of classification success. The confusion matrix, computed from the split datasets (training and testing), was used to calculate performance metrics, including accuracy, error rate, specificity, precision, recall, and F1-score. The results are shown in Table 4, which demonstrates that the ANN model effectively distinguished between Robusta coffee origins. The combination of ANN with PCA, PLS, and LDA produced different classification accuracies, with LDA achieving the highest accuracy (100%), while PCA and PLS reached <98 and <97%, respectively.

The ROC curve and AUC value were used to further evaluate the model's performance and assess its ability to differentiate between sample groups. The ROC curve is a probability curve, and the AUC represents the degree of separability between classes (Aguilar-Ruiz & Michalak, 2022).

Fig. 8 shows the ROC curves and AUC values for the combination of ANN with PCA, PLS, and LDA. The distributions in PCA and PLS showed overlaps, indicating potential errors in distinguishing between the four coffee origins. For example, the AUC for Temanggung was 0.99 in PCA and 0.98 in PLS, indicating that the model could distinguish between Temanggung and the other samples with 99% and 98% accuracy, respectively. In contrast, LDA showed no overlap between the curves, demonstrating a perfect separability of the four coffee samples, indicating that the ANN model, in combination with LDA, achieved the most accurate distinction between the positive and negative classes.

Table 4: Results of the confusion matrix and evaluation metrics from testing data

	Confusion matrix					TP	TN	FP	FN	Average (%)					
		Te	To	Da	La					AC	SP	PR	RE	FS	ER
PCA	Te	189	6	5	0	189	572	20	11	97.9	98.6	95.9	95.9	1.0	2.1
	To	6	167	0	0	167	612	7	6						
	Ma	9	0	204	1	204	573	5	10						
	La	5	1	0	199	199	586	1	6						
PLS	Te	172	8	8	0	172	576	28	16	96.2	97.5	92.5	92.4	0.9	3.8
	To	14	181	0	9	181	577	11	23						
	Ma	12	0	192	0	192	576	12	12						
	La	2	3	4	187	187	587	9	9						
LDA	Te	200	0	0	0	200	592	0	0	100	100	100	100	0.1	0.0
	To	0	192	0	0	192	600	0	0						
	Ma	0	0	204	0	204	588	0	0						
	La	0	0	0	196	196	596	0	0						

TP = true positive, TN = true negative, FP = FN = Te = Temanggung, To = Toraja, Da = Dampit, La = Lampung; AC = accuracy, SP = specificity, PR = precision, RE = recall, FS = F1-score, and ER = error rate.

Conclusion

This research demonstrates the effectiveness of combining Shortwave Near-Infrared (SWNIR) spectroscopy with various dimensional reduction techniques (PCA, PLS, and LDA) and an Artificial Neural Network (ANN) classifier to authenticate the origins of Indonesia Robusta coffee. The results revealed that the combination of LDA and ANN achieved the highest classification accuracy (100%), highlighting its superior class separability and prediction reliability.

Future studies should include a broader and more diverse set of coffee samples, incorporating additional origins and varieties to improve the generalizability and robustness of the proposed model. Expanding the spectral range and exploring advanced machine learning techniques, such as Convolutional Neural Networks (CNN), may also yield higher accuracy and further refine classification performance. Furthermore, this approach holds promise for adaptation to other agricultural products, offering potential applications in food traceability and quality assurance across various sectors.

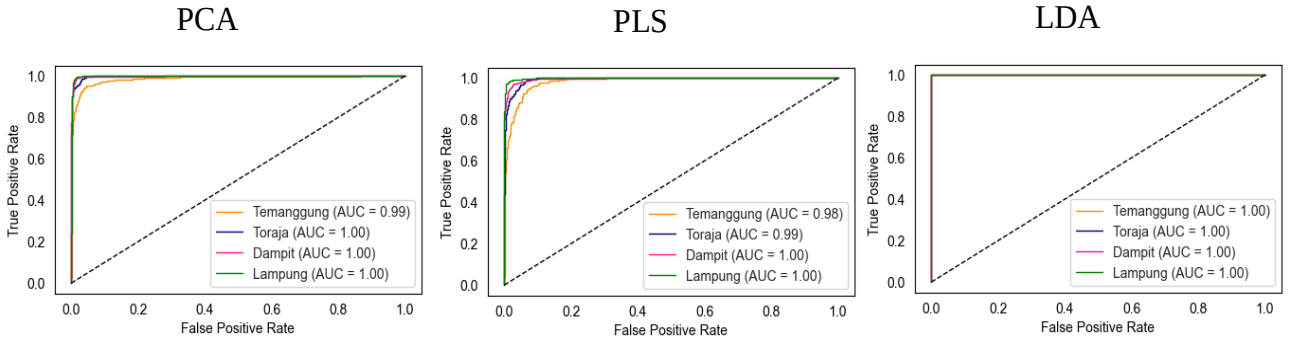


Fig. 8: The AUC and ROC curves.

DECLARATIONS

Funding: The authors received financial support from "Bantuan Pemandatan Penguatan Publikasi Internasional Bereputasi dengan Dampak Tinggi Tahun Anggaran 2023" from Universitas Gadjah Mada.

Acknowledgement: The authors thank LDPD for providing support during the study.

Conflict of Interest: The authors declare no conflicts of interest.

Data Availability: The data will be made available upon request.

Ethics Statement: This study does not involve human or animal subjects; thus, no ethical approval was required.

Author's Contribution: AD and REM planned the idea for the study; AD conducted an experiment and data analysis; REM validated the results; AD and REM wrote the manuscript draft; and REM reviewed the final draft.

Generative AI Statement: The authors declare that no Gen AI/DeepSeek was used in the writing/creation of this manuscript.

Publisher's Note: All claims stated in this article are exclusively those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated/assessed in this article or claimed by its manufacturer is not guaranteed or endorsed by the publisher/editors.

REFERENCES

Abadi, F.R., Masithoh, R.E., Sutiarso, L., & Rahayoe, S. (2024). Evaluation of Indonesian local soybean based on chemical characteristics and visible - near infrared spectra with chemometrics. *Biotropia*, 31 (1), 2054. https://doi.org/10.11598/btb.2024.31.1.2054

Abdu, N., & Mutuku, J. (2021). Willingness to pay for socially responsible products: A meta−analysis of coffee ecolabelling. *Heliyon*, 7, e07043. https://doi.org/10.1016/j.heliyon.2021.e07043

Aguilar-Ruiz, J.S., & Michalak, M. (2022). Multiclass Classification Performance Curve. *IEEE Access*, 10, 68915–68921. https://doi.org/10.1109/ACCESS.2022.3186444

Ahad, M.T., Li, Y., Song, B., & Bhuiyan, T. (2023). Comparison of CNN-based deep learning architectures for rice diseases classification. *Artificial Intelligence in Agriculture*, 9, 22–35. https://doi.org/10.1016/j.aiia.2023.07.001

Ahmed, S., Brinkley, S., Smith, E., Sela, A., Theisen, M., Thibodeau, C., Warne, T., Anderson, E., Van Dusen, N., Giuliano, P., Ionescu, K.E., & Cash, S.B. (2021). Climate Change and coffeec quality: systematic review on the effects of environmental and management variation on secondary metabolites and sensory attributes of coffea arabica and coffea canephora. *Frontiers in Plant Science*, 12, 708013. https://doi.org/10.3389/fpls.2021.708013

Ansari, R.M., Anvari, M.A., Monfared, S.K., & Gorgin, S. (2025). 4 - Linear discriminant analysis. In J. Amani Rad, S. Chakraverty, & K. B. T.-D. R. in M. L. Parand (Eds.), Dimensionality Reduction in Machine Learning (pp. 109–126). https://doi.org/https://doi.org/10.1016/B978-0-44-332818-3.00013-7

Ayesha, S., Hanif, M.K., & Talib, R. (2020). Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, 59(2019), 44–58. https://doi.org/10.1016/j.inffus.2020.01.005

Ayu, P.C., Rindang, A., & Sebayang, N.U.W. (2024). NIR Spectroscopy application to determine the carbohydrate and lipid content in green bean coffee. *IOP Conference Series: Earth and Environmental Science*, 1302(1), 7–10. https://doi.org/10.1088/1755-1315/1302/1/012114

Ban, Y., Park, H., Hong, S.J., Yu, S.Y., Moon, H.S., & Shin, E.C. (2025). Sensomics and chemometrics approaches of differential brewed and roasted coffee (Coffea arabica) from Ethiopia using biomimetic sensory-based machine perception techniques: Effects of caffeine on bitter taste and the generation of volatiles. *Food Chemistry*, 476, 143407. https://doi.org/10.1016/j.foodchem.2025.143407

Baqueta, M.R., Alves, E.A., Valderrama, P., & Pallone, J.A.L. (2023). Brazilian Canephora coffee evaluation using NIR spectroscopy and discriminant chemometric techniques. *Journal of Food Composition and Analysis*, 116, 105065. https://doi.org/10.1016/j.jfca.2022.105065

Barbin, D.F., Felicio, A.L. de S.M., Sun, D.W., Nixdorf, S.L., & Hirooka, E.Y. (2014). Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. *Food Research International*, 61, 23–32. https://doi.org/10.1016/j.foodres.2014.01.005

Bessada, S.M.F., Alves, R.C., Costa, A.S.G., Nunes, M.A., & Oliveira, M.B.P.P. (2018). Coffea canephora silverskin from different geographical origins: A comparative study. *Science of The Total Environment*, 645, 1021–1028. https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.07.201

Bhagya Raj, G.V.S., & Dash, K.K. (2022). Comprehensive study on applications of artificial neural network in food process modeling. *Critical Reviews in Food Science and Nutrition*, 62(10), 2756–2783. https://doi.org/10.1080/10408398.2020.1858398

Bowers, A.J., & Zhou, X. (2019). Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes. *Journal of Education for Students Placed at Risk*, 24(1), 20–46. https://doi.org/10.1080/10824669.2018.1523734

Brereton, R.G. (2009). Chemometrics for Pattern Recognition. In Chemometrics for Pattern Recognition. John Wiley & Sons. https://doi.org/10.1002/9780470746462

Caporaso, N., Whitworth, M.B., Grebby, S., & Fisk, I.D. (2018). Non-destructive analysis of sucrose, caffeine and trigonelline on single green coffee beans by hyperspectral imaging. *Food Research International*, 106, 193–203. https://doi.org/10.1016/j.foodres.2017.12.031

Carrington, A.M., Manuel, D.G., Fieguth, P., Ramsay, T.O., Osmani, V., Wernly, B., Bennett, C., Hawken, S., Magwood, O., Sheikh, Y., Mcinnes, M., & Holzinger, A. (2022). Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 329–341. https://doi.org/10.1109/TPAMI.2022.3145392

Chakravartula, S.S.N., Moscetti, R., Bedini, G., Nardella, M., & Massantini, R. (2022). Use of convolutional neural network (CNN) combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee. *Food Control*, 135, 108816. https://doi.org/10.1016/j.foodcont.2022.108816

Chousou, C., & Mattas, K. (2019). Assessing consumer attitudes and perceptions towards food authenticity. *British Food Journal*, 123(5), 1947–1961. https://doi.org/10.1108/BFJ-03-2019-0177

Correia, R.M., Tosato, F., Domingos, E., Rodrigues, R.R.T., Aquino, L.F.M., Filgueiras, P.R., Lacerda, V., & Romão, W. (2018). Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta*, 176, 59–68. https://doi.org/10.1016/j.talanta.2017.08.009

Davila, C., & Sîrbu, R. (2021). Determination of caffeine content in Arabica and Robusta green coffee of Indian origin. *European Journal of Natural Sciences and Medicine*, 4(1), 69–79. https://doi.org/10.26417/425qba31z

De Diego, I.M., Redondo, A.R., Fernández, R.R., Navarro, J., & Moguerza, J.M. (2022). General performance score for classification problems. *Applied Intelligence*, 52(10), 12049–12063. https://doi.org/10.1007/s10489-021-03041-7

Dharmawan, A., Masithoh, R.E., & Amanah, H.Z. (2023a). Development of PCA-MLP Model Based on Visible and Shortwave Near Infrared Spectroscopy for Authenticating Arabica Coffee Origins. *Foods*, 12(11), 2112. https://doi.org/10.3390/foods12112112

Dharmawan, A., Masithoh, R.E., & Amanah, H.Z. (2023b). Performance of A Portable NIR Spectrometer to Distinguish Coffee Species Based on Qualitative Chemometric and Artificial Neural Network (ANN) Models. *BIO Web of Conferences*, 80. https://doi.org/10.1051/bioconf/20238006007

Farhadpour, S., Warner, T.A., & Maxwell, A.E. (2024). Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices. *Remote Sensing*, 16(3), 1–22. https://doi.org/10.3390/rs16030533

Géron, A. (2019). Hands-on Machine Learning whith Scikit-Learing, Keras and Tensorfow. In O'Reilly Media, Inc. http://14.139.161.31/OddSem-0822-1122/Hands-On_Machine_Learning_with_Scikit-Learn-Keras-and-TensorFlow-2nd-Edition-Aurelien-Geron.pdf

Giraudo, A., Grassi, S., Savorani, F., Gavoci, G., Casiraghi, E., & Geobaldo, F. (2019). Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. 99, 137–145. https://doi.org/10.1016/j.foodcont.2018.12.033

Guerrero-Peña, A., Vázquez-Hernández, L., Bucio-Galindo, A., & Morales-Ramos, V. (2023). Chemical analysis and NIR spectroscopy in the determination of the origin, variety and roast time of Mexican coffee. *Heliyon*, 9(8), e18675. https://doi.org/10.1016/j.heliyon.2023.e18675

Ikram, A., Mehmood, H., Arshad, M.T., Rasheed, A., Noreen, S., & Gnedeka, K.T. (2024). Applications of artificial intelligence (AI) in managing food quality and ensuring global food security. *CyTA - Journal of Food*, 22(1), 2393287. https://doi.org/10.1080/19476337.2024.2393287

Kurniawan, F., Budiastra, I.W., Sutrisno, & Widyotomo, S. (2019). Classification of arabica Java coffee beans based on their origin using nir spectroscopy. *IOP Conference Series: Earth and Environmental Science*, 309(1), 0–8. https://doi.org/10.1088/1755-1315/309/1/012006

Liang, N., Sun, S., Zhang, C., He, Y., & Qiu, Z. (2022). Advances in infrared spectroscopy combined with artificial neural network for the authentication and traceability of food. *Critical Reviews in Food Science and Nutrition*, 62(11), 2963–2984. https://doi.org/10.1080/10408398.2020.1862045

Luna, A.S., da Silva, A.P., Alves, E.A., Rocha, R.B., Lima, I.C.A., & de Gois, J.S. (2017). Evaluation of chemometric methodologies for the classification of Coffea canephora cultivars via FT-NIR spectroscopy and direct sample analysis. *Analytical Methods*, 9(29), 4255–4260. https://doi.org/10.1039/C7AY01167A

Malta, M.R., Fassio, L. de O., Liska, G.R., Carvalho, G.R., Pereira, A.A., Botelho, C.E., Ferraz, V.P., Silva, A.D., Pedrosa, A.W., Alvaro, L.N., & Pereira, R.G.F.A. (2020). Discrimination of genotypes coffee by chemical composition of the beans: Potential markers in natural coffees. *Food Research International*, 134, 109219. https://doi.org/10.1016/j.foodres.2020.109219

Munyendo, L., Njoroge, D., & Hitzmann, B. (2022). The potential of spectroscopic techniques in coffee analysis—a review. *Processes*, 10(1), 1–25. https://doi.org/10.3390/pr10010071

Nanga, S., Bawah, A.T., Acquaye, B.A., Billa, M.-I., Baeta, F.D., Odai, N.A., Obeng, S.K., & Nsiah, A.D. (2021). Review of dimension reduction methods. *Journal of Data Analysis and Information Processing*, 09(03), 189–231. https://doi.org/10.4236/jdaip.2021.93013

Nawrocka, A., & Lamorska, J. (2016). Determination of food quality by using spectroscopic methods. In G. Stanislaw & A. Stepniewski (Eds.), *Advances in Agrophysical Research*, 11, 52722. https://doi.org/10.5772/52722

Novtahaning, D., Shah, H.A., & Kang, J.M. (2022). Deep learning ensemble-based automated and high-performing recognition of coffee leaf disease. *Agriculture (Switzerland)*, 12(11), 1909. https://doi.org/10.3390/agriculture12111909

Nunes, J.C., Ordanini, A., & Giambastiani, G. (2021). The concept of authenticity: what it means to consumers. *Journal of Marketing*, 85(4), 1–20. https://doi.org/10.1177/0022242921997081

Núñez, N., Collado, X., Martínez, C., Saurina, J., & Núñez, O. (2020). Authentication of the origin , variety and roasting fingerprinting and chemometrics. Application to the detection and quantitation of adulterated the coffee samples. *Foods*, 9(3), 378. https://doi.org/10.3390/foods9030378

Obeidat, S.M., Hammoudeh, A.Y., & Alomary, A.A. (2018). Application of FTIR Spectroscopy for assessment of green coffee beans according to their origin. *Journal of Applied Spectroscopy*, 84(6), 1051–1055. https://doi.org/10.1007/s10812-018-0585-9

Pahlawan, M.F.R., & Masithoh, R.E. (2022). Vis-NIR spectroscopy and PLS-DA model for classification of Arabica and Robusta roasted coffee bean. *Advances in Science and Technology*, 115, 45–52. https://doi.org/10.4028/p-60bbc9

Pandiselvam, R., Prithviraj, V., Manikantan, M.R., Kothakota, A., Rusu, A.V., Trif, M., & Mousavi Khaneghah, A. (2022). Recent advancements in NIR spectroscopy for assessing the quality and safety of horticultural products: A comprehensive review. *Frontiers in Nutrition*, 97, 3457. https://doi.org/10.3389/fnut.2022.973457

Poláková, K., Bobková, A., Demianová, A., Bobko, M., Lidiková, J., Jurčaga, L., Belej, Ľ., Mesárošová, A., Korčok, M., & Tóth, T. (2023). Quality attributes and sensory acceptance of different botanical coffee co-products. *Foods*, 12(14), 2675. https://doi.org/10.3390/foods12142675

Prasetyo, E.W., Amanah, H.Z., Farras, I., Pahlawan, M.F.R., & Masithoh, R.E. (2024). Partial least square regression for nondestructive determination of sucrose content of healthy and Fusarium spp. infected potato (Solanum tuberosum L.) utilizing visible and near-infrared spectroscopy. *International Journal on Advanced Science, Engineering and Information Technology*, 14(3), 19841. https://doi.org/10.18517/ijaseit.14.3.19841

Pratama, B.R., & Wisika, S.D. (2022). Individual social and economic position versus impact of collective services: Contribution to small-scale farmers' adoption on single origin coffee processing scheme. *Songklanakarin Journal of Science and Technology*, 44(1), 13–20. https://doi.org/10.14456/sjst-psu.2022.3

Priambodo, D.C., Saputro, D., Pahlawan, M.F.R., & Masithoh, R.E. (2022). Determination of Acid Level (pH) and Moisture Content of Cacao Beans at Various Fermentation Level Using Visible Near-Infrared (Vis-NIR) Spectroscopy. *IOP Conference Series: Earth and Environment*, 985, 12045. https://doi.org/10.1088/1755-1315/985/1/012045

Ribeiro, J.S., Ferreira, M.M.C., & Salva, T.J.G. (2011). Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta*, 83(5), 1352–1358. https://doi.org/10.1016/j.talanta.2010.11.001

Rismiwandira, K., Roosmayanti, F., Pahlawan, M.F.R., & Masithoh, R.E. (2020). Application of Fourier Transform Near-Infrared (FT-NIR) spectroscopy for detection of adulteration in palm sugar. *IOP Conf. Series: Earth and Environmental Science*, 012122. https://doi.org/10.1088/1755-1315/653/1/012122

Sim, J., McGoverin, C., Oey, I., Frew, R., & Kebede, B. (2024). Non-linear machine learning coupled near infrared spectroscopy enhanced model performance and insights for coffee origin traceability. *Journal of Near Infrared Spectroscopy*, *32*(4), 127-139. https://doi.org/10.1177/09670335241269014

Sornam, M., & Vanitha, V. (2018). Enhanced artificial neural network models using principal component analysis for plants multiclass classification. 2018 Tenth International Conference on Advanced Computing (ICoAC), 13–19. https://doi.org/10.1109/ICoAC44903.2018.8939105

Suhandy, D., & Yulia, M. (2018). Luwak coffee classification using UV-Vis spectroscopy data: comparison of linear discriminant analysis and support vector machine methods. *Aceh International Journal of Science and Technology*, 7(2), 115–121. https://doi.org/10.13170/aijst.7.2.8972

Torga, G.N., & Spers, E.E. (2020). Chapter 2 - Perspectives of global coffee demand. In L. F. de Almeida & E. E. B. T.-C. C. and I. S. in B. Spers (Eds.), Woodhead Publishing Series in Consumer Sci & Strat Market (pp. 21–49). https://doi.org/https://doi.org/10.1016/B978-0-12-814721-4.00002-0

Trihartono, A. (2022). Extending Indonesian gastrodiplomacy: blending geographical indication (GI) and "kopi Indonesia." *IIUM Law Journal*, 30, 69–89. https://doi.org/https://doi.org/10.31436/iiumlj.v30iS1.699

Velásquez, S., & Banchón, C. (2023). Influence of pre-and post-harvest factors on the organoleptic and physicochemical quality of coffee: a short review. *Journal of Food Science and Technology*, 60(10), 2526–2538. https://doi.org/10.1007/s13197-022-05569-z

Wahyudi, A., Wulandari, S., Aunillah, A., & Alouw, J.C. (2020). Sustainability certification as a pillar to promote Indonesian coffee competitiveness. *IOP Conference Series: Earth and Environmental Science*, 418(1), 12009. https://doi.org/10.1088/1755-1315/418/1/012009

Wang, Q., Ma, Y., Zhao, K., & Tian, Y. (2022). A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, 9(2), 187–212. https://doi.org/10.1007/s40745-020-00253-5

Wei, X., Li, S., Zhu, S., Zheng, W., Zhou, S., Wu, W., & Xie, Z. (2021). Quantitative analysis of soybean protein content by terahertz spectroscopy and chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 208, 104199. https://doi.org/10.1016/j.chemolab.2020.104199

Weyer, L.G., & Lo, S.C. (2006). Spectra - structure correlations in the near-infrared. In Handbook of Vibrational Spectroscopy. https://doi.org/10.1002/9780470027325.s4102

Xiong, Z., Cui, Y., Liu, Z., Zhao, Y., Hu, M., & Hu, J. (2020). Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. *Computational Materials Science*, 171, 109203. https://doi.org/10.1016/j.commatsci.2019.109203

Yang, J., Xu, J., Zhang, X., Wu, C., Lin, T., & Ying, Y. (2019). Deep learning for vibrational spectral analysis: Recent progress and a practical guide. *Analytica Chimica Acta*, 1081, 6–17. https://doi.org/10.1016/j.aca.2019.06.012

Yani, A., & Utomo, T.P. (2025). Chemical fingerprint of robusta coffee beans from various local clones in Tanggamus , Lampung Province , Indonesia. *Biodiversitas*, 26(5), 2222–2234. https://doi.org/10.13057/biodiv/d260520

Yusibani, E., Woodfield, P. L., Rahwanto, A., Surbakti, M.S., Rajibussalim, & Rahmi (2023). Physical and chemical properties of indonesian coffee

beans for different postharvest processing methods. *Journal of Engineering and Technological Sciences*, 55(1), 1–11. https://doi.org/10.5614/j.eng.technol.sci.2023.55.1.1

Yusmanizar, & Munawar, A.A. (2021). Environmental origin classification of coffee beans using infrared spectroscopy. *IOP Conference Series: Earth and Environmental Science*, 922(1), 12014. https://doi.org/10.1088/1755-1315/922/1/012014

Zhang, F., Yu, X., Li, L., Song, W., Dong, D., Yue, X., Chen, S., & Zeng, Q. (2025). Research on rapid and non-destructive detection of coffee powder adulteration based on portable near-infrared spectroscopy technology. *Foods*, 14(3), 536. https://doi.org/10.3390/foods14030536

Zheng, W., Fu, X., & Ying, Y. (2014). Spectroscopy-based food classification with extreme learning machine. *Chemometrics and Intelligent Laboratory Systems*, 139, 42–47. https://doi.org/10.1016/j.chemolab.2014.09.015