

Rapid Prediction of Total Gingerol Content and Harvest Age Classification of Fresh Indonesian Red Ginger Using UV-Vis Spectroscopy

Tika Hafzara Siregar^{1*}, Fareihan Rizki Farikhin², Slamet Widodo², Y. Aris Purwanto², Lamhot Parulian Manalu¹, Tri Yulni¹, Nenie Yustiningsih¹, Harianto¹, Jonni Firdaus³, Primawati Yenni Fauziah⁴, Endang Suwandi⁵, Marga Asta Jaya Mulya⁵, Astuti⁶, Tantry Eko Putri Mariastuty⁶

¹Research Center for Equipment Manufacturing Technology, National Research and Innovation Agency, South Tangerang, Banten, 15314, Indonesia

²Department of Mechanical and Biosystem Engineering, IPB University, Bogor, 16002, Indonesia

³Research Center for Horticulture, National Research and Innovation Agency, Cibinong, West Java, 16911, Indonesia

⁴Research Center for Food Technology and Processing, National Research and Innovation Agency, Yogyakarta, 55861, Indonesia

⁵Research Center for Manufacturing Technology of Production Machinery, National Research and Innovation Agency, South Tangerang, Banten, 15314, Indonesia

⁶Research Center for Processing Technology, National Research and Innovation Agency, South Tangerang, Banten, 15314, Indonesia

*Corresponding author: tika006@brin.go.id

Abstract

Red ginger (*Zingiber officinale* var. *rubrum*) is a potential spice commodity widely used in the food and pharmaceutical industries due to its high gingerol content, which has antioxidant, anti-inflammatory, and anticancer properties. This study aims to develop a predictive model for gingerol content and classification model for harvest age in Indonesia fresh red ginger using chemometrics and machine learning based on Ultraviolet-Visible (UV-Vis) spectroscopy. Samples from three harvest ages (7, 8, and 9 months after planting) were analyzed using a UV-Vis spectrometer. UV-Vis spectral data were acquired and processed using several spectral preprocessing techniques, namely baseline correction, Savitzky-Golay (SGs), Multiplicative Scatter Correction (MSC), and first derivative (D1). The analysis was performed using Partial Least Squares (PLS) and Artificial Neural Network (ANN) for prediction, as well as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), k-NN, Random Forest, and Support Vector Machine (SVM) for harvest age classification. The ANN paired with SGs+D1 with 2 hidden nodes achieved the best predictive performance ($R^2 = 0.66$, $r = 0.74$), with RMSEC and RMSEP of 0.7062 and 0.8317, respectively. Meanwhile, the SVM model with a combination of Baseline+D1 preprocessing achieved a harvest age classification accuracy of 0.91 and precision, recall, and F1-score of 0.93 each. The findings of this research yield significant practical implications for both advanced post-harvest technology and the phytopharmaceutical industry. By delivering a highly precise classification model, this work lays a solid technical foundation required to design high-throughput, automated, and cost-effective sorting systems tailored for fresh Indonesian red ginger matrices.

Keywords

Chemometrics, Classification, Gingerol, Machine Learning

Received: 17 April 2026, Accepted: 17 July 2026

<https://doi.org/10.26554/sti.2026.11.4.1406-1422>

1. INTRODUCTION

Ginger (*Zingiber officinale*) is a widely cultivated spice with a high antioxidant content, exhibiting immense potential as a primary functional ingredient in the healthcare and pharmaceutical industries (Bellik, 2014). Indonesia's national ginger production, which reaches approximately 198,873 tons per year, underscores its prominent role as an industrial commodity used extensively in refreshing beverage formulations and traditional medicine, including standardized herbal medicines and phytopharmaca (of Statistics, 2024). Red ginger possesses the highest concentration of oleoresins, essential oils, and pungency. This distinct, pungent profile is governed by volatile phe-

nolic ketones in the oleoresin matrix, specifically gingerol and shogaol. Notably, the accumulation of gingerol (23% to 25%) and shogaol (18% to 25%) in red ginger significantly exceeds levels in the gajah and emprit varieties, making it a superior pharmaceutical asset (Prasad and Tyagi, 2015).

Chemically, gingerol consists of a homologous series of phenolic ketones, with 6-gingerol standing out as the most pharmacologically active component due to its potent anti-cancer, anti-inflammatory, and antioxidant properties (Wang et al., 2014). Consequently, industrial Standard Operating Procedures (SOPs) mandate rigorous screening and verification of raw materials through a Certificate of Analysis (CoA)

to ensure compliance with physical, chemical, and microbiological quality standards before processing (Sembiring et al., 2025). In the Indonesian agricultural landscape, red ginger is systematically harvested between 7 and 9 months after planting to optimize the yield of these active secondary metabolites. However, recent studies indicate that harvest age dramatically alters the oleoresin and gingerol/shogaol equilibrium, making precise maturity determination crucial for downstream quality control and phytochemical standardization, especially for health product ingredients (Zhang et al., 2022).

Conventionally, the chemical identification and quantification of gingerol rely on destructive analytical techniques, such as High-Performance Liquid Chromatography (HPLC) and Gas Chromatography-Mass Spectrometry (GC-MS). These methods are inherently destructive, highly expensive and time-consuming, require specialized human resources, and occasionally suffer from baseline peak overlap during complex matrix identification (Ho and Su, 2015). Due to these constraints, fresh red ginger supplied by local farmers and micro-businesses frequently lacks reliable chemical quality data. This lack of real-time standardization creates a major supply chain bottleneck, resulting in a severe quality deficit in which raw materials fail to meet strict industrial pharmaceutical benchmarks.

To resolve this limitation, optical spectroscopy combined with chemometrics has emerged as a powerful non-destructive alternative for rapid agro-industrial assessment. While Near-Infrared (NIR) spectroscopy has been extensively evaluated, Ultraviolet-Visible (UV-Vis) spectroscopy offers a highly cost-effective instrumentation option with exceptional sensitivity to electronic transitions in organic phenolic structures such as gingerol (Razavi and Kenari, 2023). Nevertheless, raw UV-Vis spectral data obtained from biological tissues often suffer from pronounced baseline shifts, light scattering, and overlapping peaks, due to the complex physical-chemical matrix of fresh rhizomes. Hence, mathematical preprocessing operations such as Savitzky-Golay smoothing (SGs), Multiplicative Scatter Correction (MSC), and baseline correction are indispensable for minimizing systematic background noise and optimizing the signal-to-noise ratio (Maldonado et al., 2023; Rinnan et al., 2009). Extraction of precise qualitative and quantitative matrices from these overlapping signals further necessitates robust chemometric modeling. While linear chemometric tools such as Partial Least Squares (PLS), Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA) are widely used (Wold et al., 2001), they struggle to capture complex nonlinear variations. Conversely, Artificial Neural Networks (ANN) offer distinct machine learning advantages in handling non-linear spectral properties, potentially providing superior predictive capabilities (Massart et al., 1998; Sinaga, 2026).

Although several studies have utilized UV-Vis or NIR spectroscopy combined with PLS and ANN to evaluate phytochemical profiles in standard ginger species (Johnson et al., 2021; Subositi et al., 2022), a substantial knowledge gap persists regarding the optimization of raw materials in their fresh or wet states. Conventional spectroscopic calibrations predominantly

rely on dried or extracted samples, which remain impractical for rapid farm gate or factory reception sorting. This study distinguishes itself from the existing literature and establishes scientific novelty through three interconnected pillars.

This research focuses on a specific, authentic dataset of Indonesian red ginger (*Zingiber officinale* var. *Rubrum*) to chart the optical fingerprints of this regional variety. This work directly addresses the industrial bottleneck of predicting unstable gingerol content within a fresh, high-moisture matrix, where water absorption typically distorts absolute quantitative calibration. To address the low absolute prediction accuracy in wet conditions, we introduce a paradigm shift by leveraging rapid maturity as a robust alternative. A strategic integration of multi-class classification (LDA, SVM, RF, k-NN) and quantitative prediction (PLS, ANN) was introduced. By merging harvest age discrimination (7, 8, and 9 months) with gingerol estimation, this framework identifies the highest-accuracy pathway to develop an automated smart sorting system. Ultimately, this integrated approach provides a practical, high-throughput engineering solution to mitigate raw material quality deficits, allowing agro-industries to efficiently sort and standardize fresh red ginger based on maturity clusters before chemical processing.

2. EXPERIMENTAL SECTION

2.1 Materials

The fresh red ginger was harvested from Cianjur, West Java, Indonesia. The red ginger has a harvest age of 7, 8, and 9 months. Each harvested age consists of 30 samples, for a total of 90 fresh red ginger rhizome samples. Red Ginger is separated from the stem and then washed with water. The fresh red ginger should be stored at room temperature for a day.

2.2 Preparation of Red Ginger Extract for HPLC Analysis

The preparation of red ginger powder extract followed the method reported by Maisaroh et al. (2025) with slight modifications. 4 mg red ginger was chopped and mixed with 40 ml ethanol (Merck), then sonicated for 30 minutes. The solution was then filtered and centrifuged for 10 minutes at 10 °C using a Tomy MX-307 high-speed centrifuge to obtain the supernatant. The supernatant was dried in a vacuum oven at 40 °C. The resulting red ginger extract was stored under refrigerated conditions until analysis.

2.3 Determination of Total Gingerol

The gingerol total from fresh red ginger was determined using an ultra-high-performance liquid chromatography (UHPLC) method using a Shimadzu liquid chromatography system (UPLC Shimadzu, LC20AD, Shimadzu Co., Japan). Standard solutions of 6-gingerol, 8-gingerol, and 10-gingerol were prepared by dissolving accurately weighed portions of the standards in acetonitrile. Calibration mixtures were prepared with final concentrations of 32.5, 62.5, 125, 250, and 500 µg/mL. Separation was performed on a C18 reversed-phase column using a mobile phase composed of 0.1% formic acid (A) and

acetonitrile (B). The flow rate was 1 mL/min, the injection volume was 20 μ l, and the column temperature was maintained at 25 °C (Guo et al., 2019). UV absorbance was monitored between 200 and 400 nm, and with quantification at 280 nm. For the determination of gingerol content, the red ginger extract was diluted to 2000 ppm, filtered through a PTFE hydrophilic syringe filter (0.22 μ m pore size, 13 mm diameter), and transferred into a 2 mL vial.

2.4 Moisture Content

The moisture content method described in the Indonesian National Standard (SNI) 01-2891:1992. A total of 5 g of fresh Indonesian red ginger was accurately weighed and placed in a pre-weighed container. The sample was dried in an oven at 105 °C for 24 hours, then cooled in a desiccator for 15 minutes and reweighed. The process of heating in the oven, cooling in the desiccator, and reweighing was repeated until a constant weight was obtained. The moisture content was calculated based on the difference in weight before and after drying using the following equation:

$$\text{Moisture content (\%)} = \frac{W1}{W} \times 100\% \quad (1)$$

W = The sample weight before drying (g)

W1 = The weight loss after drying (g)

2.5 Data Spectral Collection

Fresh Indonesian red ginger spectrum collected based on harvest ages of 7, 8, and 9 months. Each harvest consisted of 30 samples. The samples were placed in a UV-Vis Spectrometer (Ocean Optics Maya Pro 2000) over the wavelength range of 200-900 nm, with a spectral resolution of 1 nm. Spectral were acquired in triplicate, yielding 90 spectral per harvesting age.

2.6 Preprocessing Data Spectral

Data processing of spectral data is very important to minimize interference from wave interference and noise that appears in the measurement results. Spectral data processing is performed using MATLAB R2018a (MathWorks Inc., USA). The preprocessing methods used include Savitzky-Golay smoothing (SGs), baseline, MSC, and their combinations with first derivatives (Yu et al., 2022). The Savitzky-Golay algorithm fits a polynomial to successive segments of the spectrum, thereby replacing the original data points with smoothed values that preserve the overall shape and key spectral features (CAMO, 2015). Essentially, SGs are employed to reduce random high-frequency instrumental noise present in the spectral without significantly distorting signal characteristics. In this study, SGs are particularly critical because fresh red ginger samples contain complex cellular structures that can cause minor high-frequency fluctuations in UV-Vis light reflectance, which must be eliminated to reveal the true chemical absorption of gingerol (Castillejos-Mijangos et al., 2022).

Multiplicative scatter correction (MSC) was utilized to correct spectral variations arising from differences in particle size and other physical properties of the samples. This method effectively minimizes undesired multiplicative and additive effects, such as signal amplification and baseline shifts, thereby enhancing spectral comparability (Munir et al., 2021). For fresh red ginger, MSC is indispensable because of the rhizome's uneven surface slices and fibrous texture, which naturally introduce variations in optical path length and light scattering. By applying these methods, spectral differences caused by physical sample variations are decoupled from the chemical variations of gingerol content (Rinnan et al., 2009; Naskar et al., 2025).

Baseline correction plays a crucial role in transforming spectral that initially exhibit non-uniform absorbance offsets. By aligning the spectral's starting positions along the vertical axis, this method produces a more consistent and comparable dataset. In UV-Vis measurements of fresh agricultural products, baseline drift often occurs due to sample moisture variability and slight positioning inconsistencies at the spectrometer window. Baseline correction ensures that these external optical offsets do not interfere with the quantitative modeling of gingerol (Eilers and Boelens, 2005).

Furthermore, combinations of the first derivative with the aforementioned preprocessing techniques were also investigated. The application of the first derivative enhances spectral resolution by reducing baseline drift and sharpening spectral peaks, enabling clearer identification of relevant spectral features associated with the chemical constituents of fresh red ginger. Since the UV-Vis spectrum of gingerol often overlaps heavily with other phenolic compounds and pigments (such as anthocyanins, which give red ginger its distinctive color), the first derivative is used to resolve these overlapping peaks. This isolation allows the machine learning models (ANN and LDA) to capture specific wavelengths highly correlated with gingerol and harvest maturity (Pasquini, 2003; Cen and He, 2007).

2.7 Classification Model for Harvesting Age

In this study, Principal Component Analysis (PCA) was performed using Matlab R2018a as an initial, unsupervised exploratory tool to reduce data complexity and visualize the distribution patterns of the high-dimensional spectral dataset. This statistical technique operates by transforming hundreds or thousands of highly correlated spectral variables into a significantly smaller set of uncorrelated variables, known as principal components (PCs), typically fewer than ten, by identifying the factors that capture the largest sources of variation in the dataset (Park et al., 2003). Each PC is a linear combination of the original variables, where the first principal component (PC₁) accounts for the highest proportion of variance, and subsequent components (PC₂, PC₃, etc.) are constructed to be orthogonal to one another, with the total explained variance systematically decreasing in the order of PC₁ > PC₂ > PC₃ (Granato et al., 2018). Beyond effective dimensionality reduction, the primary merit of employing PCA in this research lies in its capacity as an unbiased exploratory method that does not incorporate class

label information, thereby allowing for an objective evaluation of the inherent data structures and the natural tendency of the red ginger samples to cluster according to their respective harvest ages (7, 8, and 9 months).

Following the exploratory PCA analysis, supervised classification models were developed to accurately categorize fresh Indonesian red ginger samples by harvest age (7, 8, and 9 months). Linear Discriminant Analysis (LDA) was employed as the baseline classification method, which maximizes the ratio of between-class variance to within-class variance to achieve optimal class separation (Qu and Pei, 2024). To evaluate and potentially enhance classification performance against complex non-linear spectral patterns, the performance of LDA was compared with three prominent machine learning algorithms: Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), and Random Forest (RF). To ensure optimal classification accuracy, the hyperparameters for each machine learning model were carefully tuned as follows (Table 1).

Table 1. Hyperparameters for Different Algorithms Using Machine Learning

Model	Parameter	Value
SVM	Kernel	RBF
	Kernel Scale	Auto
	CV	5 Fold
	Standardization	Z-score
KNN	K	5
	Distance	Euclidean
	CV	5 Fold
RF	Number of Trees	200
	OOB	On
	CV	5 Fold

2.8 Prediction Model for Gingerol Content

The gingerol content prediction models in this study were developed and compared using Partial Least Squares (PLS) regression and an Artificial Neural Network (ANN). PLS is a robust linear regression approach that is highly suitable for calibration with a relatively small number of samples, effectively handling the high dimensionality and collinearity of combined chemical and spectroscopy datasets (Williams and Norris, 1987). However, to address the potential non-linear relationships between the UV-Vis spectral and gingerol concentration, an ANN model was constructed.

The ANN model developed in this research utilized a supervised learning method trained with the backpropagation algorithm. The backpropagation algorithm is highly versatile and widely applicable for multilayer networks that incorporate different activation functions under a supervised learning paradigm (Rumelhart et al., 1986). To establish robust calibration models, minimize data selection bias, and ensure a reliable evaluation of both linear and non-linear algorithms, a 5-fold

cross-validation technique was systematically implemented for both methods. Under this validation scheme, the entire dataset-comprising UV-Vis spectral data over 220-450 nm and the corresponding reference gingerol content-was randomly partitioned into five equal-sized subsamples. In successive iterations, four folds were used as the training set to calibrate the models. In contrast, the remaining fold was reserved as the independent validation set for testing, allowing every sample to be used for both training and testing across five replication cycles.

For the linear approach, the PLS regression model optimized the latent variables within each fold to handle the high dimensionality of the spectral data. Concurrently, the non-linear ANN architecture was constructed using three sequential layers: an input layer (spectral wavelengths), a hidden layer, and an output layer (predicted gingerol content). The network training process within each fold was executed in MATLAB R2018a (MathWorks Inc., USA) and configured as a feed-forward multi-layer perceptron (MLP) trained with the Levenberg-Marquardt backpropagation algorithm for a maximum of 1,000 epochs and 1,000 iterations.

The LM algorithm was selected due to its fast convergence rate and high efficiency in minimizing errors in spectroscopic non-linear mapping (Hagan and Menhaj, 1994). Statistical parameters, including the coefficient of determination (R^2), root mean square error of calibration (RMSEC), and root mean square error of prediction (RMSEP), were calculated across the folds to assess the performance and predictive capability of each model. The best predictive framework was determined by the highest R^2 value, the lowest prediction errors (RMSEC and RMSEP), and the model’s overall stability, offering a fast, robust, and non-destructive method for monitoring gingerol levels using UV-Vis spectroscopy.

3. RESULTS AND DISCUSSIONS

3.1 Gingerol Content of Red Ginger in Different Harvesting Age

Indonesian red ginger contains 6-gingerol, 8-gingerol, and 10-gingerol. These chemical compounds are useful in the health sector, exhibiting anti-inflammatory (Yücel et al., 2022), antioxidant (Alharbi et al., 2022), antivirulence (Fareid et al., 2025), anticancer (Abdull Rahim et al., 2023), antimicrobial (Fareid et al., 2025), and antidiabetic effect (Rahmawati et al., 2022). Among its bioactive constituents, 6-gingerol is considered the major compound contributing to its antibacterial (Elfaky et al., 2024), antioxidant (Ahmed et al., 2021), and anticancer properties (Wala et al., 2022), highlighting the potential of red ginger as a valuable raw material for the health and pharmaceutical industries.

The determination of gingerol content in red ginger samples using High-Performance Liquid Chromatography (HPLC) is highly effective because 6-gingerol, as the major bioactive compound, is generally more readily detected due to its relatively higher concentration compared to other minor constituents, which may be present at lower levels or exhibit weaker detectability within the HPLC system. The HPLC method

is universally recognized as a robust tool for the separation, identification, and quantitative analysis of organic compounds from complex herbal and agricultural matrices (Gusti et al., 2021; Riwu et al., 2023; Salamah et al., 2025).

In regional agro-industrial research, validated HPLC methods have been widely used to establish rigorous baselines for natural antioxidant profiles and to quantify target secondary metabolites from Indonesian endemic flora (Rahmawati et al., 2023). Furthermore, optimizing such chromatographic protocols remains essential to ensure that complex plant-based compounds are accurately resolved and mapped before implementing high-throughput spectral predictions (Cahyani et al., 2023). Figure 1 presents the HPLC analysis results of fresh Indonesian red ginger, showing the presence of several chemical compounds.

Table 2. Identification of Major Compounds in Fresh Indonesian Red Ginger

RT (min)	Name of the Compound	Molecular Formula
4.884	6-Gingerol	C ₁₇ H ₂₆ O ₄
9.437	8-Gingerol	C ₁₉ H ₃₀ O ₄
11.619	10-Gingerol	C ₂₁ H ₃₄ O ₄
20.811	6-Shogaol	C ₁₇ HO ₃

Based on HPLC analysis of fresh red ginger, the chromatograms in Figure 1 confirm the successful separation and identification of the target gingerol homologs. The compounds were well resolved with distinct retention times (RT): 6-gingerol at 4.884 min, 8-gingerol at 9.437 min, and 10-gingerol at 11.619 min (Table 2). From the peak area integrations, 6-gingerol was confirmed as the most dominant compound, followed by 8-gingerol and 10-gingerol, indicating its strong potential as a major source of functional ingredients for pharmaceutical raw materials.

In Indonesia, the pharmaceutical industry requires red ginger raw materials to contain at least 1.5% gingerol total content. Therefore, the recommended harvest age for red ginger intended for medicinal raw materials is at least 9 months to ensure compliance with industrial quality standards. This was further supported by the present study, which showed that the highest gingerol content was found in ginger harvested at 9 months (2.56 ± 0.1%), whereas ginger harvested at 7 months contained only 1.13 ± 0.48% gingerol. The measured gingerol content in fresh red ginger samples is presented in Table 3.

In general, the concentrations of all three gingerol compounds increased with increasing harvest age, with samples harvested age of 9 months showing the highest levels among the evaluated groups. This trend indicates that plant maturation plays a significant role in the biosynthesis and accumulation of bioactive compounds, particularly gingerols. In addition to gingerol content, moisture content was also measured as a complementary parameter in the analysis. This is important

Table 3. The Gingerol Total Content in Fresh Indonesian Red Ginger

Harvest Age	Gingerol Total Content (%)
7 Month	1.13 ± 0.48
8 Month	1.23 ± 0.89
9 Month	2.56 ± 0.11

because moisture content can profoundly influence optical absorbance measurements, which are highly sensitive to water molecules and hydrogen bonding.

High moisture levels in fresh red ginger can significantly distort spectral characteristics and, consequently, compromise the accuracy of quantitative analytical models. In the field of Indonesian agricultural and food chemometrics, managing the complex interplay among plant matrices, moisture variations, and target bioactive compounds remains a central challenge for ensuring prediction reliability (Gusti et al., 2021). Furthermore, investigating these physical-chemical interactions is essential because fluctuations in matrix composition can cause baseline shifts that mask underlying organic peaks (Yudha S. et al., 2024). The moisture content values of fresh red ginger are presented in Table 4. Based on the moisture content measurements of the 90 samples, no significant differences were observed among red ginger harvested at 7, 8, and 9 months.

3.2 Uv-Vis Spectral Characteristic

UV-Vis spectral are generated when incident light interacts with a sample and is absorbed at specific wavelengths corresponding to electronic transitions within the molecule, in which electrons move from one energy level to another (Skoog et al., 1996). The interpretation of UV-Vis spectra is based on understanding how molecular electrons absorb energy and transition between orbitals. These transitions may include π → π*, n → π*, or charge transfer processes, depending on the chemical structure and functional groups present in the compounds (Hollas, 2004). Figure 2 illustrates the original UV-Vis spectral of fresh Indonesian red ginger within the wavelength range of 200-900 nm.

The chemical constituents of the analyzed material influence the presence of peaks and valleys in a spectral. In addition to chemical factors, spectral features are influenced by the sample's physical properties, such as particle size and surface characteristics (Blanco and Villarroya, 2002). Based on the average absorbance spectral shown in Figure 2, the absorbance intensity within the 500-900 nm wavelength range appears relatively stable, with minimal variation across samples. This is evidenced by the closely overlapping spectral lines corresponding to different harvesting ages, particularly between red ginger harvested at 7 and 8 months. This indicates that this region is less informative and lacks sufficient sensitivity to discriminate gingerol content or differentiate samples based on harvest age.

In contrast, the wavelength range of 200-220 nm exhibits considerable noise, as indicated by sharp, irregular fluctuations

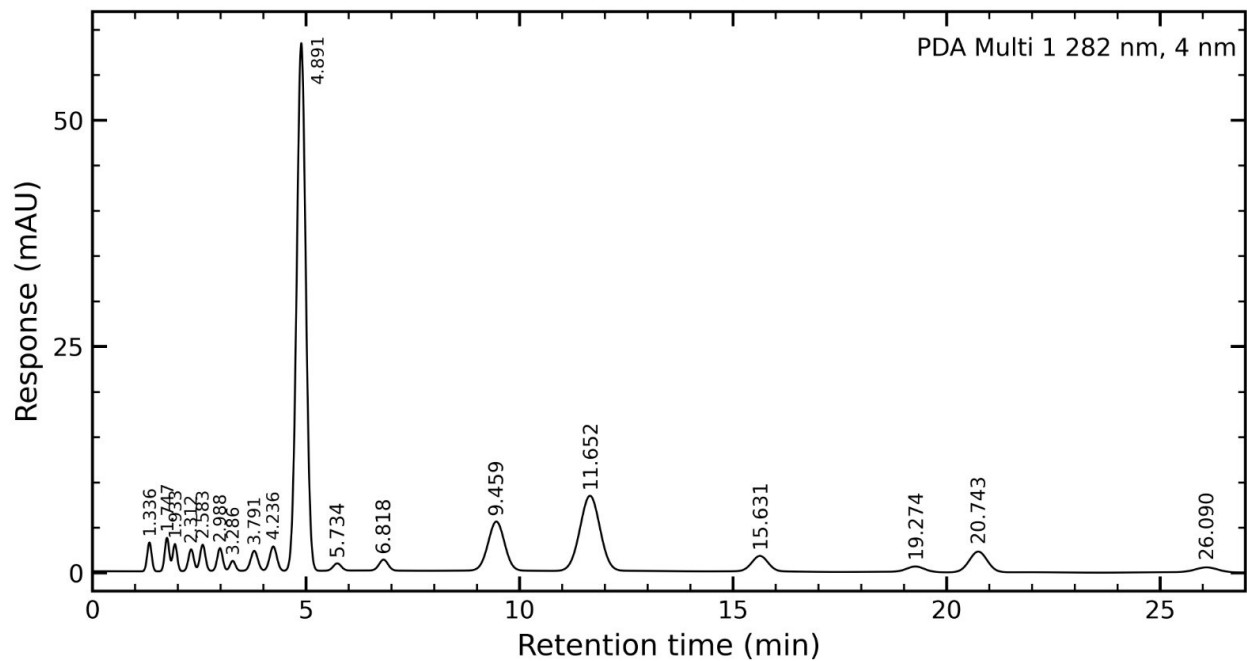


Figure 1. HPLC Chromatogram of Chemical Compounds in Indonesian Red Ginger

Table 4. The Moisture Content of Fresh Indonesian Red Ginger

Harvest Age (Month)	Moisture Content (%)		
	Value	Minimum	Maximum
7	86.01 ± 4.59	76.05	92.23
8	85.09 ± 5.13	74.55	94.25
9	85.24 ± 5.20	77.04	94.34

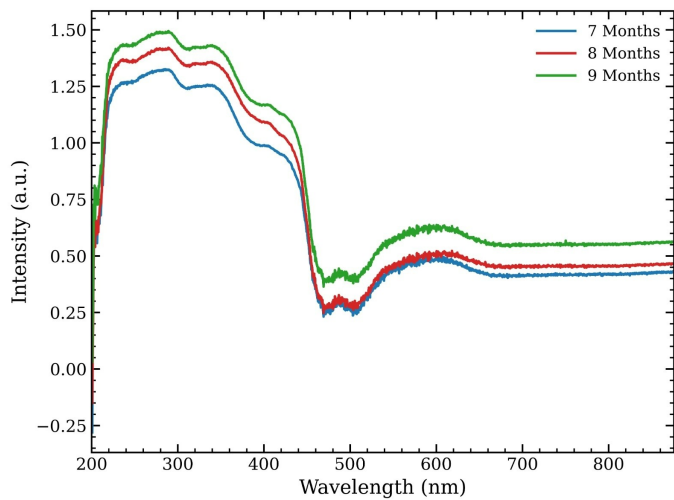


Figure 2. Original UV-Vis Absorbance Spectral in the Range of 200-900 nm

in the spectral curves. This instability may arise from instrumental limitations and strong absorption of solvent or matrix components at low wavelengths. Therefore, this region was excluded from further analysis. Subsequent analysis was focused on the wavelength range of 220-450 nm, where more distinct and chemically relevant absorbance features are observed. This region is particularly important for compounds containing conjugated systems, such as phenolic compounds and gingerol derivatives.

Gingerols, including 6-gingerol, possess aromatic rings with conjugated double bonds and carbonyl groups, which contribute to characteristic electronic transitions. The absorption bands observed in the range of approximately 270-300 nm are commonly associated with $\pi \rightarrow \pi^*$ transitions of the aromatic ring system, while weaker bands at higher wavelengths may be attributed to $n \rightarrow \pi^*$ transitions involving carbonyl functional groups (Skoog et al., 1996; Hollas, 2004). These spectral characteristics are consistent with the presence of phenolic structures and support the use of this wavelength region for analytical modeling. Furthermore, variations in absorbance intensity within this region may reflect differences in gingerol

concentration across samples with different harvesting ages. Therefore, the selected range of 220-450 nm provides more relevant chemical information and improves the robustness of subsequent chemometric analysis for predicting gingerol content and classifying red ginger samples. The original UV-Vis spectral within this selected wavelength range (220-450 nm) are presented in Figure 3.

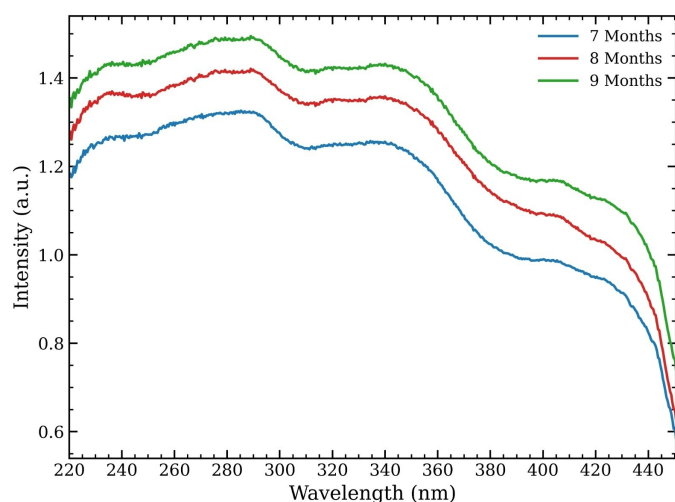


Figure 3. Original UV-Vis Absorbance Spectral in the Range of 220-450 nm

Based on the average UV-Vis absorbance spectral within the wavelength range of 220-450 nm shown in Figure 3, several prominent absorbance peaks are observed, particularly at 220-230 nm and 280-290 nm. Clear differences in absorbance intensity among harvesting ages are evident at these wavelengths, indicating that these regions are highly informative for analytical modeling. The observed variation in absorbance follows the linear relationship described by the Beer-Lambert law, suggesting that changes in spectral intensity are directly associated with differences in gingerol concentration with harvest age. In general, the spectral intensity increases with increasing harvesting age of red ginger. This observation is consistent with the Beer-Lambert law, as described by Mäntele and Deniz (2017), which states that the absorbance of a solution is directly proportional to the concentration of the absorbing species and the optical path length.

These findings are consistent with previous studies. Sukweenadhi et al. (2023) reported strong absorption at 200 nm, 224 nm, 226 nm, and 282 nm for standard mixtures of gingerol-related compounds. Meanwhile, You et al. (2019) demonstrated that 230 nm provides higher sensitivity and superior peak resolution compared to 280/282 nm, which are commonly used reference wavelengths according to ISO standards.

Furthermore, the separation of spectral patterns among samples harvested age at 7, 8, and 9 months supports their utility in classification models. Techniques such as Principal Component Analysis (PCA) or supervised classification algo-

rithms are likely to leverage these wavelength regions as key discriminative variables, enabling effective differentiation based on harvest age. As a complementary validation, the comparative graph of average gingerol total content across harvest ages (Figure 4) demonstrates a consistent trend with the observed spectral variations. This alignment between chemical reference data and spectral features further confirms the reliability of the selected wavelength range (220-450 nm) for both quantitative prediction and classification.

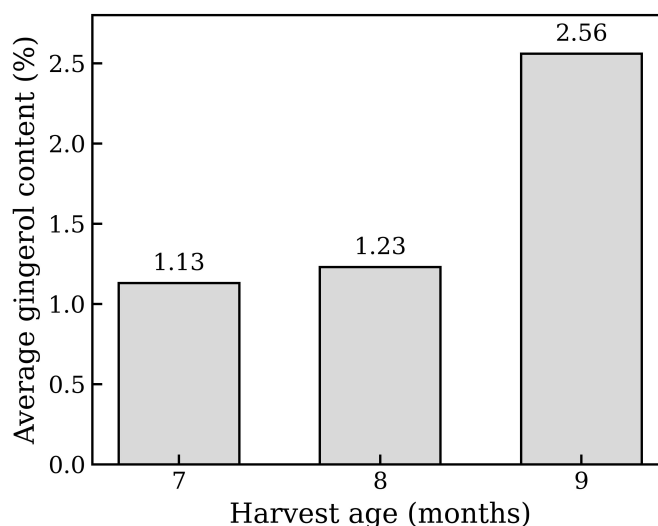


Figure 4. Comparison of the Average Gingerol Total Content Based on the Harvesting Age of Fresh Indonesian Red Ginger

Furthermore, HPLC analysis revealed a similar trend: an increase in the total content of the three major gingerol compounds with increasing harvest age. Specifically, red ginger harvested at 9 months exhibited the highest average gingerol content compared to samples harvested at earlier stages, as presented in Figure 4. The consistency between the UV-Vis spectral patterns and the HPLC results indicates that the observed increase in absorbance intensity is strongly associated with the elevated concentration of gingerol compounds. This finding confirms that the Beer-Lambert relationship applies to UV-Vis spectral analysis of fresh red ginger, supporting the reliability of this technique for monitoring changes in chemical composition associated with harvesting age.

3.3 Preprocess UV-Vis Spectral

The application of preprocess methods illustrate in Figure 5. The application of SGs results in smoother spectral profiles by reducing high-frequency noise (Figure 5a). However, no substantial differences are observed between the original and SGs-treated spectral in overall spectral patterns, indicating that this method primarily enhances signal quality without significantly altering spectral features.

Additional preprocessing methods, including baseline correction and multiplicative scatter correction (MSC), were also

applied. As shown in Figure 5b, baseline correction produces a noticeable visual shift in the spectral, particularly at the lower portion of the curves. Spectral that initially exhibited non-uniform absorbance offsets become more aligned along the vertical axis, resulting in a more consistent baseline across samples.

MSC was employed to correct for baseline variations and spectral distortions caused by differences in measurement conditions or instrumental fluctuations. As shown in Figure 5c, MSC improves spectral uniformity, reducing intensity variability compared to the raw data. This method operates by adjusting each spectral to a common reference, effectively minimizing multiplicative scattering effects while preserving the original absorbance scale. Visually, MSC stabilizes spectral patterns across samples, thereby facilitating more reliable chemical interpretation.

Further preprocessing involved combining SGs, baseline correction, and MSC with the first derivative (D1), as shown in Figures 5d, 5e, and 5f, respectively. The incorporation of the first derivative enhances spectral resolution by reducing baseline drift and sharpening spectral peaks, allowing for clearer identification of relevant spectral features.

Overall, these combined preprocessing approaches improve the visibility of significant spectral characteristics, stabilize signal variation, and enhance the performance of subsequent multivariate analysis. As shown in Figures 5d-5f, the resulting spectra are more structured, with sharper, more distinct peaks. These improvements indicate that the preprocessing combinations effectively reduce noise and baseline interference while highlighting subtle spectral differences among samples.

3.4 Principal Component Analysis (PCA) of Red Ginger Based on Harvesting Age

Spectral data were analyzed using Principal Component Analysis (PCA) to qualitatively evaluate the distribution and clustering patterns of the spectral dataset. The UV-Vis spectral range employed in this study was 220-450 nm, a region previously identified as the most informative for analysis. In this study, PCA was applied as an exploratory tool to facilitate visual interpretation of spectral groupings based on different treatments. PCA reduces the dimensionality of the dataset while preserving the most significant sources of variation, enabling the identification of natural clustering patterns. This method provides an efficient representation of the overall dataset by extracting principal components that capture the maximum variance within the spectral data (Park et al., 2003). Before PCA, spectral were corrected or preprocessed to minimize the effects of noise and wavelength interference in the raw spectral data. This step is essential to improve the accuracy and stability of the analysis, ensuring that the extracted components reflect meaningful chemical variation rather than instrumental or environmental artifacts (Siregar et al., 2018). Figure 6 presents the PCA results obtained from the original UV-Vis spectral.

The PCA results for UV-Vis spectral based on the original data did not exhibit clear or significant clustering among the

different harvest age groups of red ginger. Therefore, a further PCA analysis was performed on the preprocessed spectral data to enhance discrimination capability. To enhance the visibility of these characteristic absorption features, first derivative (D1) preprocessing was applied in combination with previously implemented methods, including Savitzky-Golay smoothing (SGs+D1), baseline correction (baseline+D1), and multiplicative scatter correction (MSC+D1). The application of the first derivative improves spectral resolution by emphasizing subtle peak variations that may not be clearly observable in the original spectral, thereby enhancing the detection of chemically relevant information associated with gingerol compounds. The PCA score plots of the UV-Vis spectral after these preprocessing combinations are presented in Figure 7 (a-c). All three preprocessing approaches demonstrate a similar pattern of separation, with a more distinct clustering trend than the results from the original and singly preprocessed spectral. This indicates that incorporating the first derivative significantly improves PCA's ability to capture meaningful variation related to harvesting age.

However, the separation between samples harvested at ages 8 and 9 months remains suboptimal, as indicated by the noticeable overlap in the PCA score plots. This overlap suggests that the chemical composition, particularly the gingerol content, between these two harvesting stages is relatively similar. According to the Beer-Lambert law, this observation implies that the absorbance values-and thus the concentration of gingerol-do not differ substantially between harvesting ages of 8 and 9, resulting in limited spectral differentiation. Overall, the results indicate that derivative-based preprocessing enhances spectral discrimination, although further advanced modeling approaches may be required to achieve complete classification between closely related harvest stages.

3.5 Classification Models Performance

The classification was performed using spectral data in the 220-450 nm wavelength range in the UV-Vis region, which was identified as the most informative range for analysis. Based on the classification model evaluations of the UV-Vis spectral of red ginger, it was observed that the combination of the classification algorithm and the spectral data pretreatment method strongly influenced model performance. Table 5 compares performance metrics for machine learning model classification. The LDA model built using the original spectral achieved an accuracy of 0.76, with relatively balanced precision 0.76, recall 0.75, and F1-score 0.75. This indicates that the raw spectral data already contain relevant information for classification; however, performance remains moderate due to noise and spectral variability. The combination of baseline correction with the first derivative (baseline+D1) restored model performance to a higher level (accuracy = 0.80), comparable to that of MSC alone. This indicates that derivative preprocessing is more effective when applied to properly baseline-corrected spectral. The best classification performance was achieved with MSC combined with the first derivative (MSC+D1), yielding

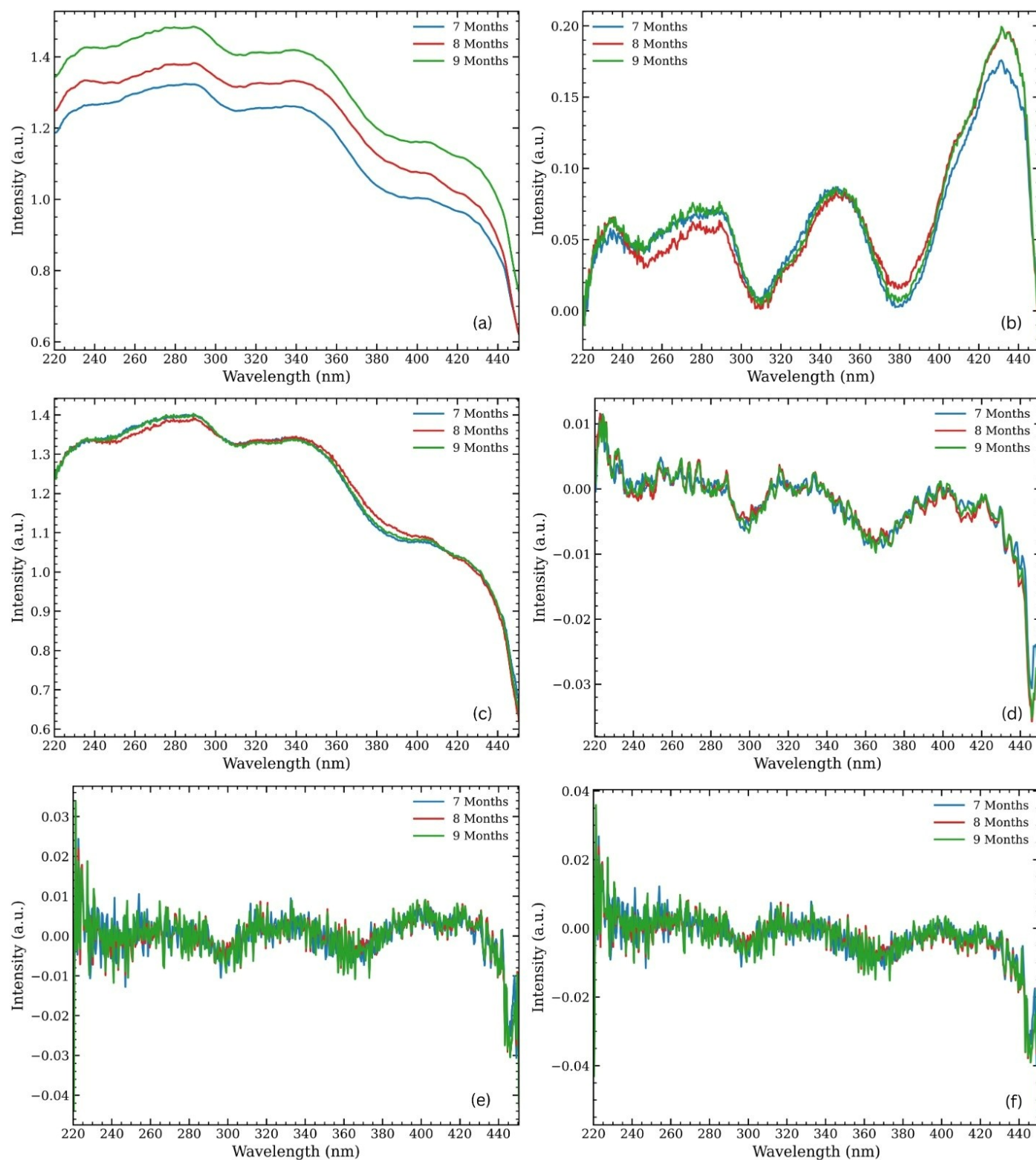


Figure 5. Average UV-Vis Spectral After Preprocessing: (a) SGs; (b) Baseline Correction; (c) MSC; (d) SGs+D1; (e) Baseline+D1; (f) MSC+D1

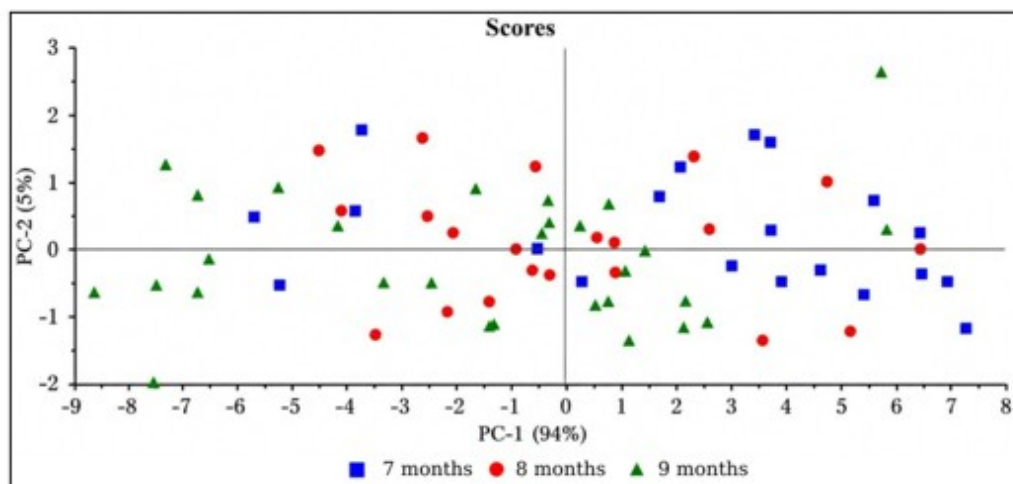


Figure 6. PCA Plot of UV-Vis Original Spectral

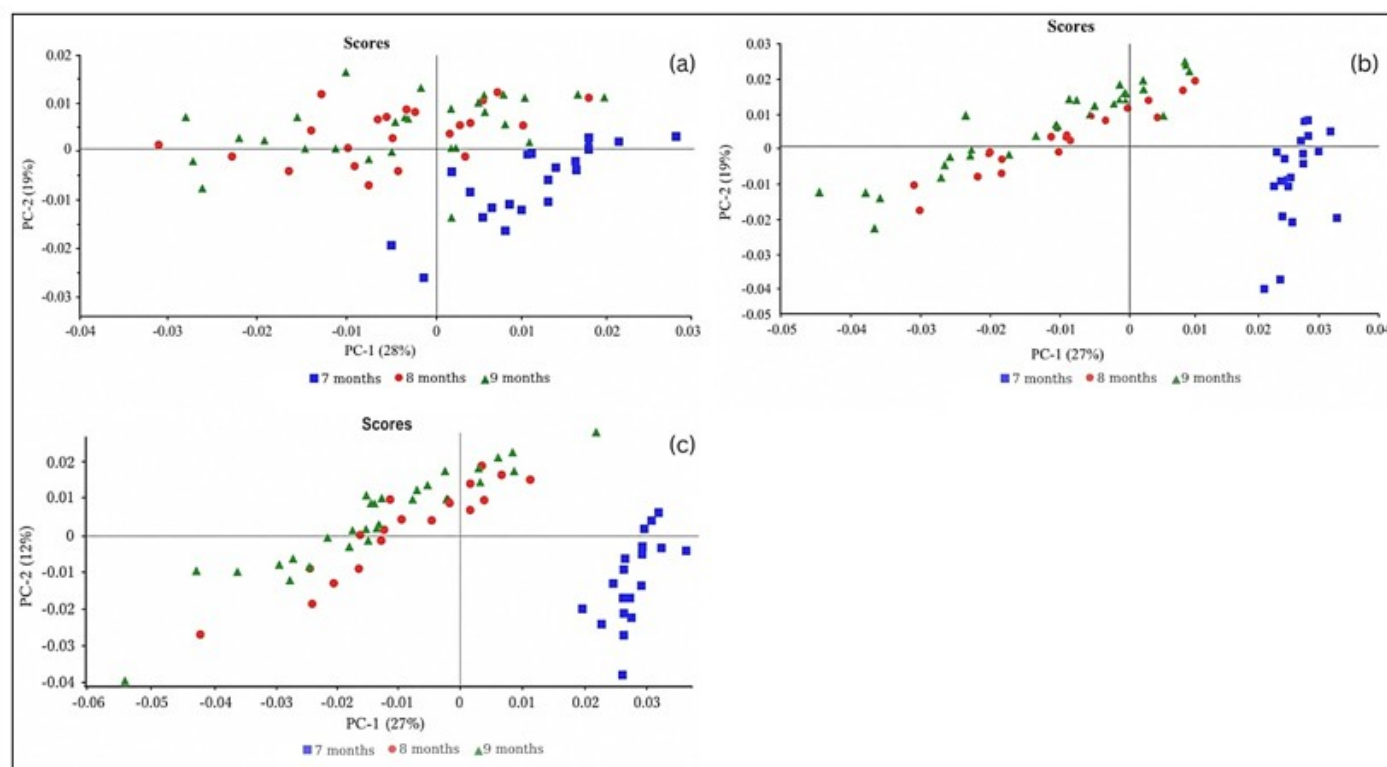


Figure 7. PCA Score Plots of UV-Vis Spectral After Preprocessing: (a) SGs+D1; (b) Baseline+D1; (c) MSC+D1

the highest accuracy, precision, recall, and F1-score (all equal to 0.90). This result demonstrates that the combination of scatter correction and derivative enhancement provides optimal conditions for highlighting subtle spectral differences associated with harvesting age. The MSC step effectively minimizes non-chemical variability, while the first derivative enhances resolution and emphasizes relevant spectral features associated with gingerol compounds.

On the other hand, in the raw (original) dataset without any preprocessing, the model performances were relatively poor; the accuracy for Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN) was only 0.33, while Random Forest (RF) achieved a marginally higher accuracy of 0.47. This deficient performance across all evaluation parameters—including accuracy, precision, recall, and F1-score—indicates that raw UV-Vis spectral lack the robust discriminatory power

needed to differentiate red ginger samples by harvest age. This limitation is primarily due to non-informative spectral variations, such as baseline shifts, instrumental noise, variations in absorbance intensity, and highly overlapping absorption bands among samples (Rinnan et al., 2009). Consequently, crucial chemical information related to physiological changes during harvesting maturity could not be effectively extracted by the classification models in their initial state.

The application of Savitzky–Golay smoothing (SGs) yielded only limited improvements, raising the SVM accuracy to 0.43, while k-NN and RF achieved 0.32 and 0.41, respectively. This implies that smoothing alone is insufficient to enhance the discrimination power of the models regarding harvest age. Although SGs effectively suppress localized random high-frequency instrumental noise, it fails to eliminate systematic baseline drift and light scattering variations that mask underlying chemical markers (Cen and He, 2007). Conversely, baseline correction provided a more pronounced enhancement, with both SVM and RF attaining an accuracy of 0.66. This improvement underscores the vital role of baseline alignment in stabilizing spectral patterns, thereby allowing the algorithms to discern maturity-related differences. Meanwhile, the Multiplicative Scatter Correction (MSC) pretreatment delivered a moderate performance enhancement on SVM (accuracy = 0.64; F1-score = 0.65), yet it failed to outperform baseline correction, particularly when applied to RF and k-NN. This indicates that MSC-based scatter correction was not the dominant factor in improving UV-Vis spectral classification for this fresh red ginger dataset. The most substantial performance breakthrough was achieved when preprocessing techniques were coupled with the first derivative (D1).

The combination of SGs + D1 elevated the accuracies of SVM, k-NN, and RF to 0.88, 0.76, and 0.84, respectively. Similarly, the MSC + D1 combination demonstrated high efficacy, achieving classification accuracies of 0.89 with SVM and 0.86 with RF. However, the optimal overall synergy was achieved by combining Baseline + D1 preprocessing with the SVM classifier. Under this configuration, the SVM model achieved an accuracy of 0.91, a precision of 0.93, a recall of 0.91, and an F1-score of 0.91. These higher metrics indicate that the model not only achieves excellent true-positive rates but also maintains an exceptional balance between identifying true class memberships and minimizing cross-class misclassifications across harvest ages.

The comparative analysis demonstrates that while the linear LDA model provides a solid and stable baseline classification—achieving its peak accuracy of 0.90 under the MSC + D1 framework—it is ultimately outperformed by advanced machine learning approaches when coupled with optimal spectral preprocessing. The raw, non-preprocessed datasets fail universally across all algorithms due to severe spectral overlaps, baseline drifts, and physical scattering inherent in fresh rhizome matrices. However, once these non-informative variances are systematically suppressed, the non-linear classification capabilities of machine learning algorithms become highly pronounced.

Baseline + D1 achieved the absolute pinnacle of classification synergy for the SVM model, with an outstanding accuracy of 0.91 and precision, recall, and F1-score of 0.93 each. This superior performance underscores the capacity of SVM to construct more precise, high-dimensional separating hyperplanes compared to the linear boundaries of LDA, particularly when handling complex biochemical transitions during the late maturation stages of red ginger. Consequently, while LDA with scatter-derivative preprocessing remains a viable option, integrating baseline correction and first-derivative mathematical enhancement with an SVM classifier is highly recommended as the most robust and accurate non-destructive framework for predicting harvesting age via UV-Vis spectroscopy.

3.6 PLS Model for Total Gingerol Prediction

The calibration and validation results of the PLS models for UV-Vis spectral, including those with various preprocessing treatments, are summarized in Table 6. Meanwhile, Figure 8 presents the calibration and validation plots for the best-performing preprocessing method, illustrating the relationship between predicted and reference gingerol values.

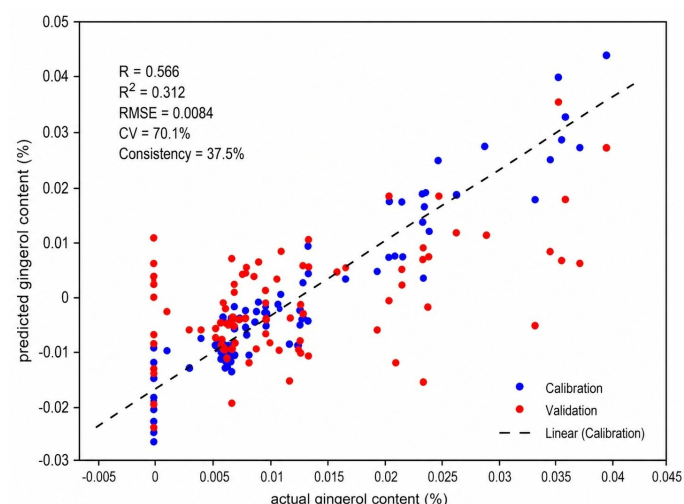


Figure 8. Calibration and Validation Results of PLS Model for Gingerol Prediction

Based on the results presented in Table 6, the statistical parameters obtained from the calibration and validation of the UV-Vis spectral data, both original and preprocessed using Partial Least Squares (PLS) regression, indicate that the linear models do not meet the optimal criteria commonly required for reliable quantification. The coefficient of determination (R^2) values across all treatments remains significantly below the desirable agro-industrial threshold (≤ 0.80), suggesting that the developed linear frameworks exhibit limited overall predictive capacity.

Among the evaluated preprocessing techniques, the Baseline correction method yielded the best performance, achieving a correlation coefficient (r) of 0.57 and a coefficient of determination (R^2) of 0.31. This was closely followed by the Baseline

Table 5. Comparison of Statistical Performance Metrics for Classification Models

Preprocess Data	Parameter	LDA	SVM	KNN	RF
Original	Accuracy	0.76	0.33	0.33	0.47
	Precision	0.76	0.11	0.11	0.47
	Recall	0.75	0.33	0.33	0.47
	F1 Score	0.75	0.17	0.17	0.46
SGs	Accuracy	0.71	0.43	0.32	0.41
	Precision	0.72	0.43	0.32	0.41
	Recall	0.72	0.43	0.32	0.41
	F1 Score	0.70	0.43	0.32	0.41
Baseline	Accuracy	0.71	0.66	0.57	0.66
	Precision	0.72	0.66	0.53	0.66
	Recall	0.72	0.66	0.57	0.66
	F1 Score	0.72	0.66	0.53	0.66
MSC	Accuracy	0.80	0.64	0.53	0.53
	Precision	0.81	0.66	0.48	0.52
	Recall	0.81	0.64	0.53	0.53
	F1 Score	0.81	0.65	0.48	0.53
SGs+D1	Accuracy	0.59	0.88	0.76	0.84
	Precision	0.72	0.88	0.78	0.84
	Recall	0.62	0.88	0.76	0.84
	F1 Score	0.60	0.88	0.73	0.84
Baseline+D1	Accuracy	0.80	0.91	0.78	0.87
	Precision	0.81	0.93	0.87	0.88
	Recall	0.81	0.91	0.78	0.87
	F1 Score	0.81	0.91	0.75	0.87
MSC+D1	Accuracy	0.90	0.89	0.70	0.86
	Precision	0.90	0.89	0.83	0.86
	Recall	0.90	0.89	0.70	0.86
	F1 Score	0.90	0.89	0.62	0.86

combined with first derivative (Baseline + D1), which achieved an R^2 of 0.26 ($r = 0.51$, $LV = 4$). In contrast, applying SGs alone severely degraded the linear model, resulting in the lowest predictive capacity ($R^2 = 0.04$, $r = 0.35$), even though it exhibited a high mathematical consistency of 80.16%.

To determine the most viable model, priority is typically given to architectures with higher Ratio of Prediction to Deviation (RPD) values, higher r values, and robust consistency parameters (ideally between 80% and 110%). Furthermore, a minimal discrepancy between the Root Mean Square Error of Calibration (RMSEC) and the Root Mean Square Error of Prediction (RMSEP) indicates strong precision and low geometric bias (Williams and Norris, 1987). Under these considerations, while the SGs model demonstrated high consistency (80.16%), its extraordinarily low validation capacity ($R^2 = 0.04$) makes it functionally obsolete.

Consequently, the Baseline-preprocessed PLS model with 12 latent variables represents the best linear compromise for predicting gingerol content in fresh red ginger, yielding RM-

SEC = 0.0032%, RMSEP = 0.0084%, RPD = 1.21, CV = 70.06%, and a consistency of 37.53%. Nonetheless, these optimized values still fall short of the requirements for robust quantitative screening. The validation R^2 of 0.31 indicates that the linear model accounts for only 31% of the biochemical variability within the samples, while the high CV (70.06%) suggests substantial prediction dispersion. Finally, the maximum achieved RPD of 1.21 reinforces that the linear PLS approach remains inadequate for high-precision absolute quantification, as values below 1.5 are typically considered insufficient for reliable non-destructive prediction in fresh agricultural matrices.

The calibration and validation plots shown in Figure 8 further support these findings. The data points are widely scattered around the regression line, which ideally should follow a 45° slope. A prominent structural anomaly is observed near the origin ($x = 0$), where a severe vertical clustering of data points spans from negative prediction values up to approximately 0.02%. This distinct phenomenon strongly indicates that at low gingerol concentrations, the linear PLS framework cannot

Table 6. Comparison of Statistical Parameters from PLS Model

Preprocess	LV	r	R ²	RMSEC (%)	RMSEP (%)	CV (%)	RPD	Consistency (%)
Original	12	0.49	0.18	0.0036	0.0092	76.31	1.11	39.62
SGs	9	0.35	0.04	0.0080	0.0099	82.78	1.03	80.16
Baseline	12	0.57	0.31	0.0032	0.0084	70.06	1.21	37.53
MSC	12	0.51	0.21	0.0025	0.0090	74.88	1.13	27.63
SGs + D1	8	0.48	0.16	0.0042	0.0093	77.55	1.10	44.90
Baseline + D1	4	0.51	0.26	0.0032	0.0087	72.69	1.17	37.18
MSC + D1	6	0.47	0.21	0.0021	0.0090	75.25	1.13	23.58

successfully resolve true biochemical absorption peaks from the overarching spectral noise and water-matrix interference inherent in fresh, wet biological tissues.

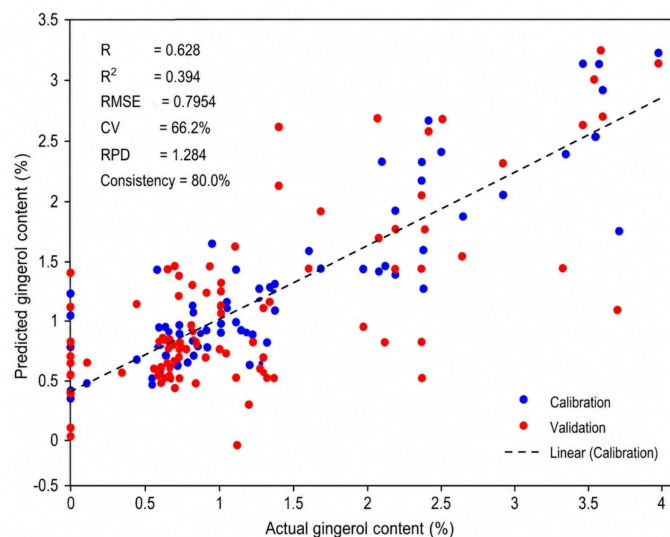
For instance, a comparative benchmark utilizing Near-Infrared Spectroscopy (NIRS) on dried ginger for gingerol total prediction achieved a robust cross-validation capability with an $R^2 > 0.70$ and RPD of approximately 2 via PLSR calibration (Johnson et al., 2021). This massive discrepancy is directly attributable to the state of the plant matrix; while the referenced NIRS framework benefited from dried, crushed powder where moisture variables were completely eradicated, our short-wavelength UV-Vis PLS configuration was directly subjected to fresh ginger. The dominant water-matrix absorption bands and severe baseline drifts inherent in un-dried tissues completely disrupt linear correlation vectors, proving that standard linear combinations of latent variables are fundamentally incapable of extracting weak gingerol signatures from highly fluid biological matrices. This physical limitation further strengthens the empirical requirement to bypass linear calibrations entirely in favor of non-linear algorithms when screening fresh agricultural commodities.

Furthermore, as the actual gingerol content increases, the validation data points (represented in red) diverge progressively from the calibration baseline, with pronounced underestimation and overestimation. This severe scattering pattern provides diagnostic evidence of model underfitting, indicating that a strict linear combination of latent variables cannot adequately capture the complex, non-linear chemical interactions within the fresh red ginger matrix. Overall, these visual profiles conclusively demonstrate that the UV-Vis based PLS model, irrespective of the baseline or derivative preprocessing steps applied, lacks the necessary mathematical flexibility for precise quantification. Consequently, capturing these complex spectrochemical trends demands a transition toward more advanced nonlinear modeling frameworks, such as artificial neural networks, or the implementation of targeted variable selection algorithms to eliminate uninformative wavelengths.

3.7 ANN Model for Total Gingerol Prediction

The absorbance spectral obtained from UV-Vis spectroscopy, previously preprocessed and analyzed using PLS, were further evaluated using an Artificial Neural Network (ANN) approach.

Unlike PLS, which models linear relationships, an ANN was employed to capture potential nonlinear relationships between gingerol concentration and UV-Vis spectral absorbance data. The calibration and validation results of the ANN models based on UV-Vis spectral data are presented in Table 7. Meanwhile, Figure 9 illustrates the performance of the best ANN model, showing the relationship between predicted and reference gingerol values during model evaluation.

**Figure 9.** Calibration and Validation Results of ANN Model for Gingerol Prediction

Based on the results presented in Table 7, the statistical parameters obtained from the calibration and validation of the UV-Vis spectral data utilizing the non-linear ANN framework demonstrate a noticeable shift in data-mapping behavior compared to the linear PLS approach. Across all evaluated preprocessing regimes, the network architectures optimized their hidden-layer capacities using a remarkably small number of hidden nodes, typically 1-5.

The ANN model paired with SGs+D1, with 2 hidden nodes, achieved the best mathematical fit, yielding a correlation coefficient (r) of 0.74 and a coefficient of determination (R^2) of 0.66. This indicates that the network's non-linear trans-

Table 7. Comparison of Statistical Parameters from ANN Model

	Node	r	R ²	RMSEC	RMSEP	RPD	CV	Consistency
Original	4	0.71	0.54	0.7320	0.7766	1.29	68.73	91.37
SGs	5	0.66	0.51	0.7248	0.7760	1.27	66.44	96.50
Baseline	2	0.63	0.48	0.7520	0.7919	1.31	68.38	95.30
MSC	2	0.72	0.56	0.6783	0.7556	1.36	67.07	90.72
SGsD1	2	0.74	0.66	0.7062	0.8317	1.27	71.43	92.58
BaselineD1	1	0.70	0.60	0.7229	0.7844	1.27	68.81	92.42
MSCD1	2	0.66	0.52	0.9473	0.9118	1.10	80.25	105.63

fer functions captured up to 66% of the biochemical variance in gingerol levels. Other preprocessing methods also showed moderate validation fits, such as the Multiplicative Scatter Correction (MSC) with 2 nodes ($r = 0.72$, $R^2 = 0.56$) and the Baseline+D1 with 1 node ($r = 0.70$, $R^2 = 0.60$).

However, despite the apparent improvement in linear correlation and R^2 scores, a critical look at the error profiles and predictive standard metrics reveals severe architectural constraints. The absolute error parameters magnified significantly, with the best model (SGs+D1) exhibiting an RMSEC of 0.7062 and an RMSEP of 0.8317. This inflation is further reflected in the high Coefficients of Variation (CVs), which range from 66.44% to 80.25%, with the SGs+D1 model showing a CV of 71.43%. Such high CV thresholds indicate a substantial spread in prediction errors relative to the actual reference concentrations, fundamentally compromising the network’s analytical precision.

The Ratio of Performance to Deviation (RPD) values across all models remained extremely low, ranging from 1.10 (MSC+D1) to 1.36 (MSC). For the highest-fitting SGs+D1 framework, the RPD was bounded at 1.27. This diagnostic constraint is highly consistent with previous chemometric benchmarks on high-moisture agricultural matrices; for instance, [Yu et al. \(2022\)](#) similarly reported that mapping raw bioactive concentrations in ginger roots via conventional spectroscopic setups frequently restricts RPD thresholds below 1.50 due to intense water absorption bands scattering light.

Even though all models maintained excellent mathematical stability-demonstrated by highly robust consistency parameters well within the ideal range of 90.72% to 105.63% the low RPD ceiling definitively disqualifies these networks from use as standalone absolute quantification tools. Because standard chemometric conventions dictate that an RPD below 1.5 indicates inadequate predictive performance for agricultural matrices, these results confirm that while non-linear mapping enhances pattern recognition, it remains insufficient to overcome the compounding spectroscopic noise caused by the high-water content and complex physical matrices of fresh red ginger.

Further insights can be drawn from the calibration and validation scatter plots of the optimized non-linear model shown in [Figure 9](#). Based on the validation parameters embedded in the best architecture (hidden node = 2), the framework achieved

RVal values of 0.628 and 0.394. It means that the ANN successfully accounts for approximately 39.4% of the absolute variance in gingerol content. Visually, compared to the previous linear PLS models, the non-linear ANN demonstrates improved capability to align data points toward the ideal 45° regression line, particularly at higher concentrations (>2.0%).

However, the data points still exhibit considerable dispersion around the calibration line, resulting in a substantial validation error (RMSEVal = 0.7954) and a high coefficient of variation (CV Val = 66.2%). A persistent challenge is observed at the origin ($x = 0$), where a vertical array of predicted values spans from 0% up to nearly 2.2%. This structural scatter confirms that even with non-linear mapping capabilities, the neural network struggles to differentiate true low-level gingerol absorption bands from the strong background noise and water chemical matrices typical of undried, fresh red ginger.

Furthermore, the model yielded a low consistency index of 65.0% and a restricted Ratio of Performance to Deviation (RPDVal) of 1.284. While the nearly identical scattering patterns between the calibration (blue) and validation (red) datasets confirm that the system is free from overfitting, the low consistency and suppressed RPD ceiling indicate underfitting. This suggests that a simple three-layer feed-forward network with only two hidden nodes remains structurally limited when faced with complex biological matrices.

To highlight the profound impact of matrix complexity on chemometric reliability, our moderate predictive results were compared with previous spectroscopic investigations on more uniform organic matrices. In a study targeting sucrose content estimation using UV-Vis spectral absorbance, the application of Support Vector Regression (SVR) and Partial Least Squares Regression (PLSR) achieved exceptional validation metrics, with the SVR model yielding an R^2 of 0.98, a very low RMSE (0.97), and an extraordinary RPD of 5.80 ([Razavi and Kenari, 2023](#)). While sucrose in honey or liquid solutions presents a highly uniform, highly soluble target with distinct, isolated spectral signatures, the gingerol content in fresh red ginger is locked within an un-dried cellular structure. The overlapping electronic transitions of gingerol homologs are easily obscured by the shifting baseline scattering of raw plant fibers and dominant ambient water bands. Consequently, while a clear chemical matrix allows linear or simple machine learn-

ing models to reach high-precision quantification thresholds ($RPD > 5$), the raw, multi-layered red ginger matrix heavily dampens spectral resolution, verifying why our UV-Vis setup functions optimally as a qualitative sorting framework rather than an absolute quantitative assay.

The presence of these prominent residual errors and low concentration outliers heavily dampens the R^2 scores, demonstrating that further refinement is mandatory before UV-Vis spectroscopy can be deployed for dependable, non-destructive quantitative screening. These findings indicate that while the ANN model demonstrates improved capability in capturing non-linear relationships compared to linear approaches, its predictive performance based on UV-Vis spectral remains insufficient and requires further optimization, such as feature selection, data augmentation, or advanced network tuning.

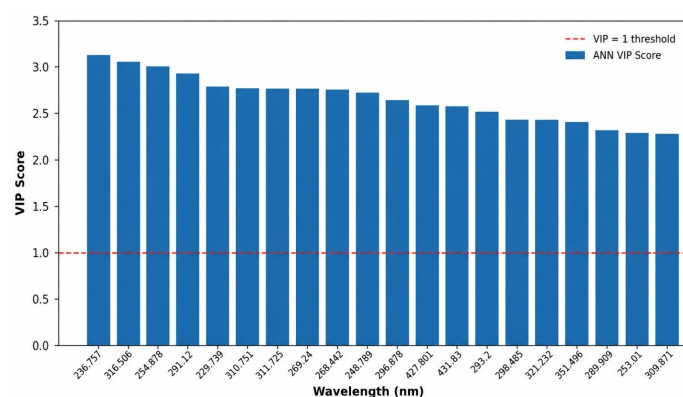


Figure 10. Variable Importance in Projection (VIP) Score of ANN Model

Variable Importance in Projection (VIP) analysis was performed on the optimized ANN model. As illustrated in Figure 10, the top 20 selected wavelengths exhibited exceptionally high VIP scores ranging from 2.30 to 3.13. The most critical wavelengths identified by the network were centered at 236.757 nm, 314.506 nm, 254.978 nm, and 287.120 nm. These wavelengths are expected to exhibit high regression coefficients and high variable importance in projection (VIP) scores, indicating strong contributions to predicting total gingerol content. The sensitivity of these regions to both aromatic and carbonyl chromophores enhances the robustness and predictive performance of the model. Importantly, these chemically relevant wavelength regions also play a critical role in chemometric modeling.

4. CONCLUSIONS

This study successfully evaluated the use of UV-Vis spectroscopy combined with chemometrics as a non-destructive tool for monitoring the quality parameters of fresh Indonesian red ginger according to harvest age. The raw spectral characteristics of the fresh, wet rhizome matrix revealed major absorbance bands in the ultraviolet region (220–450 nm), heavily dominated by complex, overlapping electronic transitions from gingerol

homologs, alongside prominent baseline shifts and moisture-induced light scattering typical of undried biological tissues. The best classification model using SVM as a supervised classification method was obtained from the preprocessed spectral data using Baseline+D1, with accuracy = 0.91, precision = 0.93, f1-score = 0.91, and recall = 0.91. For quantitative estimation, the ANN paired with SGs+D1 with 2 hidden nodes delivered the best predictive capability ($R^2 = 0.66$, $r = 0.74$). However, its high prediction errors (RMSEP = 0.8317) and limited RPD ceiling (≤ 1.36) indicate that it remains constrained for absolute high-precision quantification due to water-matrix interferences. Crucially, this research offers a substantial practical contribution to the development of post-harvest agricultural technology and the phytopharmaceutical industry. The highly accurate classification model developed in this work provides a solid scientific foundation for the design of rapid, automated, and cost-effective sorting systems for fresh Indonesian red ginger raw materials. Instead of relying on destructive, time-consuming chemical assays, industrial processing plants can deploy this spectroscopic framework at the reception gate to fast-track screening.

5. ACKNOWLEDGEMENT

The authors would like to acknowledge the Indonesia Endowment Fund for Education Agency (LPDP), Ministry of Finance, and National Research and Innovation Agency (BRIN) Indonesia, for funding this research according to RIIM-7 Grant No. 61/II.7/HK/2024.

REFERENCES

- Abdull Rahim, U., N. A. Che Roos, M. Mustapa, A. Nurdin, N. M. Taridi, and Y. A. Mohd Yusof (2023). Anticancer Effects of Gingerol, Shogaol and Curcumin in Cervical Cancer: A Systematic Review Protocol. *Advances in Human Biology*, **13**(3); 246–251
- Ahmed, S. H. H., T. Gonda, and A. Hunyadi (2021). Medicinal Chemistry Inspired by Ginger: Exploring the Chemical Space around 6-Gingerol. *RSC Advances*, **11**(43); 26687–26699
- Alharbi, K. S., M. S. Nadeem, O. Afzal, S. I. Alzarea, A. S. Altamimi, W. H. Almalki, B. Mubeen, S. Iftikhar, L. Shah, and I. Kazmi (2022). Gingerol, a Natural Antioxidant, Attenuates Hyperglycemia and Downstream Complications. *Metabolites*, **12**(12); 1274
- Bellik, Y. (2014). Total Antioxidant Activity and Antimicrobial Potency of the Essential Oil and Oleoresin of *Zingiber officinale* Roscoe. *Asian Pacific Journal of Tropical Disease*, **4**(1); 40–44
- Blanco, M. and I. Villarroya (2002). NIR Spectroscopy: A Rapid-Response Analytical Tool. *Trends in Analytical Chemistry*, **21**; 240–250
- Cahyani, I. M., Adhyatmika, E. Lukitaningsih, and T. N. S. Sulaiman (2023). Development and Validation of the RP-HPLC Method for Dexamethasone Sodium Phosphate De-

- termination in Nasal Chitosan Microsphere Preparations. *Science and Technology Indonesia*, **8**(4); 636–644
- CAMO (2015). *The Unscrambler Method Reference*. CAMO Software AS
- Castillejos-Mijangos, L. A., A. Acosta-Caudillo, T. Gallardo-Velázquez, G. Osorio-Revilla, and C. Jiménez-Martínez (2022). Uses of FT-MIR Spectroscopy and Multivariate Analysis in Quality Control of Coffee, Cocoa, and Commercially Important Spices. *Foods*, **11**(4); 579
- Cen, H. and Y. He (2007). Theory and Application of Near Infrared Reflectance Spectroscopy in Determination of Food Quality. *Trends in Food Science & Technology*, **18**(2); 72–83
- Eilers, P. H. and H. F. Boelens (2005). Baseline Correction with Asymmetric Least Squares Smoothing. *Leiden university medical centre report*, **1**(1); 5
- Elfaky, M. A., H. M. Okairy, H. M. Abdallah, A. E. Koshak, G. A. Mohamed, S. R. Ibrahim, A. A. Alzain, W. A. Hegazy, E.-S. Khafagy, and N. M. Seleem (2024). Assessing the Antibacterial Potential of 6-Gingerol: Combined Experimental and Computational Approaches. *Saudi Pharmaceutical Journal*, **32**(5); 102041
- Fareid, M. A., G. M. El-Sherbiny, N. M. Elafandy, N. E. El-toum, M. S. Othman, M. H. Sharaf, A. M. Shehabeldine, E. E. Farghal, and F. A. Hamada (2025). Shogaol from *Zingiber officinale*: Antimicrobial, Anti-Virulence, and Therapeutic Potential against Foodborne Pathogens and Beyond. *Progress in Biomaterials*, **14**(01)
- Granato, D., J. S. Santos, G. B. Escher, B. L. Ferreira, and R. M. Maggio (2018). Use of Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) for Multivariate Association between Bioactive Compounds and Functional Properties in Foods: A Critical Perspective. *Trends in Food Science & Technology*, **72**; 83–90
- Guo, F., C. Yang, C. Zang, Y. Shang, B. Zhang, H. Yu, C. Li, and S. Duan (2019). Comparison of the Quality of Chinese Ginger Juice Powders Prepared by Different Drying Methods. *Journal of Food Process Engineering*, **42**(7); e13252
- Gusti, N., D. Oktarina, R. Elvia, E. Nursa'adah, R. W. Wardhana, A. Sudaryono, and M. L. Firdaus (2021). Facile Detection of Oil Adulteration Using UV-Visible Spectroscopy Coupled with Chemometrics Analysis. *Science and Technology Indonesia*, **6**(1); 14–18
- Hagan, M. T. and M. B. Menhaj (1994). Training Feedforward Networks with the Marquardt Algorithm. *IEEE Transactions on Neural Networks*, **5**(6); 989–993
- Ho, S. C. and M. S. Su (2015). Optimized Heat Treatment Enhances the Anti-Inflammatory Capacity of Ginger. *International Journal of Food Properties*, **19**(8); 1884–1898
- Hollas, J. M. (2004). *Modern Spectroscopy*. John Wiley & Sons
- Johnson, J. B., J. S. Mani, K. B. Walsh, and M. Naiker (2021). Measurement of Gingerols and 6-Shogaol in Ginger Using Near-Infrared Spectroscopy. In *The International Conference on NIR*. Springer, pages 81–90
- Maisaroh, P. Atmaji, Astuti, W. B. Setianto, E. Restiawaty, and Y. Bindar (2025). Headspace Gas Chromatography/Mass Spectroscopy Analysis of Volatile Organic Compounds in Red Ginger (*Zingiber officinale* var. *Rubrum*): Effects of Drying on Chemical and Functional Properties. In *IOP Conference Series: Earth and Environmental Science*, volume 1477. IOP Publishing, page 012068
- Maldonado, L. M., K. Marcinkevicius, A. R. Alvarez, V. M. Salomon, M. Gaggiotti, N. Messina, M. A. Palacio, S. Medina, M. M. Maldonado, and E. Orellano (2023). UV-Vis Spectroscopy and Chemometric: A Simple Method to Differentiate Honey, Partially Ripe Honey and Nectars. *RIA. Revista de Investigaciones Agropecuarias*, **49**(3); 110–118
- Mäntele, W. and E. Deniz (2017). UV-VIS Absorption Spectroscopy: Lambert-Beer Reloaded. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **173**; 965–968
- Massart, D. L., B. G. Vandeginste, L. M. C. Buydens, P. J. Lewi, J. Smeyers-Verbeke, and S. de Jong (1998). *Handbook of Chemometrics and Qualimetrics*. Elsevier Science Inc.
- Munir, N., M. Nugent, D. Whitaker, and M. McAfee (2021). Machine Learning for Process Monitoring and Control of Hot-Melt Extrusion: Current State of the Art and Future Directions. *Pharmaceutics*, **13**(9); 1432
- Naskar, S., D. Sing, S. Banerjee, A. Shcherbakova, A. Bandyopadhyay, A. Kar, P. K. Haldar, N. Sharma, P. K. Mukherjee, and R. Bandyopadhyay (2025). Rapid Quality Assessment and Traceability of Ginger Powder from Northeast India and Indian Market Based on Near Infrared Spectroscopic Fingerprinting. *Phytochemical Analysis*, **36**(2); 369–376
- of Statistics, B. C. B. (2024). Production of Biofarmaka (Medicinal) Crops, 2021–2023
- Park, B., J. A. Abbott, K.-J. Lee, C. H. Choi, and K. H. Choi (2003). Near-Infrared Diffuse Reflectance for Quantitative and Qualitative Measurement of Soluble Solids and Firmness of Delicious and Gala Apples. *Transactions of the ASAE*, **46**(6); 1721–1731
- Pasquini, C. (2003). Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *Journal of the Brazilian Chemical Society*, **14**(2); 198–219
- Prasad, S. and A. K. Tyagi (2015). Ginger and Its Constituents: Role in Prevention and Treatment of Gastrointestinal Cancer. *Gastroenterology Research and Practice*, **2015**(1); 142979
- Qu, L. and Y. Pei (2024). A Comprehensive Review on Discriminant Analysis for Addressing Challenges of Class-Level Limitations, Small Sample Size, and Robustness. *Processes*, **12**(7); 1382
- Rahmawati, F. Y., Y. R. Utami, M. A. Zahra, and I. Maryati (2022). Naturopathic Therapy Efforts: Potential Gingerol Content in Ginger (*Zingiber officinale*) As An Antidiabetic. *Journal of Nursing Care*, **5**(1); 50–58
- Rahmawati, L., S. Widodo, D. P. Kurniadi, P. Daud, A. Triyono, Sriharti, N. D. Susanti, N. K. I. Mayasti, A. Indriati, L. E. Yulianti, et al. (2023). Determination of Colorant Type in Yellow Tofu Using Vis-NIR and SW-NIR Spectroscopy. *Food Science and Technology*, **43**; e112422
- Razavi, R. and R. E. Kenari (2023). Ultraviolet-Visible Spectroscopy Combined with Machine Learning As a Rapid De-

- tection Method to the Predict Adulteration of Honey. *Helvion*, **9**(10); e20973
- Rinnan, Å., F. van den Berg, and S. B. Engelsen (2009). Review of the Most Common Pre-Processing Techniques for Near-Infrared Spectra. *TrAC – Trends in Analytical Chemistry*, **28**(10); 1201–1222
- Riwu, A. G., J. Nugraha, D. A. Purwanto, and E. A. Triyono (2023). Determination of (+)-Catechin and Antioxidant Activity in Faloak (*Sterculia quadrifida* R. Br.) Stem Bark Infusion. *Science and Technology Indonesia*, **8**(1); 59–65
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). Learning Representations by Back-Propagating Errors. *Nature*, **323**(6088); 533–536
- Salamah, N., R. A. Firdaus, A. Guntarti, M. Ahda, A. H. Uddin, and M. F. Kurniawan (2025). Analysis of Hydrolyzed Collagen in Facial Serum for Halal Authentication Using FTIR Spectroscopy and Multivariate Calibration. *Science and Technology Indonesia*, **10**(4); 1270–1279
- Sembiring, B. B., S. Suhirman, C. Winarti, H. Haris, T. H. Siregar, E. Maryana, N. Harimurti, F. Manoi, and I. Agustinisari (2025). Physical and Chemical Properties of Rat Taro Simplicia From Various Drying Methods. *The Scientific World Journal*, **2025**(1); 6239186
- Sinaga, B. (2026). Crop Yield Prediction Using Artificial Neural Network with Principal Component Analysis (PCA) Dimensionality Reduction. *Journal of Information Technology and Computer Science*, **11**(1); 34–45
- Siregar, T., U. Ahmad, Sutrisno, and A. Maddu (2018). Fluorescence Spectroscopy Characteristics of Indonesian Citrus. In *Journal of Physics: Conference Series*, volume 1057. IOP Publishing, page 012011
- Skoog, D. A., D. M. West, F. J. Holler, S. R. Crouch, et al. (1996). *Fundamentals of Analytical Chemistry*, volume 33. Saunders College Pub. Fort Worth
- Subositi, D., H. Widodo, R. Mujahid, N. Rahmawati, F. Mustofa, M. Indrian, S. Haryanti, I. Sholikhah, A. Maruzy, and Y. Widiyastuti (2022). Genetic Diversity and Chemicals Profile of Ginger (*Zingiber officinale* Roscoe) in Indonesia. *Int J Adv Sci Eng Inform Technol*, **12**(3); 929–936
- Sukweenadhi, J., P. D. Damitasari, K. Kartini, P. Christanti, and E. N. Putri (2023). Gingerol and Shogaol on Red Ginger Rhizome (*Zingiber officinale* var. *rubrum*) Using High-Performance Liquid Chromatography. *Pharmaciana*, **13**(2); 166
- Wala, K., W. Szlasa, N. Sauer, P. Kasperkiewicz-Wasilewska, A. Szewczyk, J. Saczko, N. Rembiałkowska, J. Kulbacka, and D. Baczyńska (2022). Anticancer Efficacy of 6-Gingerol with Paclitaxel against Wild Type of Human Breast Adenocarcinoma. *Molecules*, **27**(9)
- Wang, S., C. Zhang, G. Yang, and Y. Yang (2014). Biological Properties of 6-Gingerol: A Brief Review. *Natural Product Communications*, **9**(7); 1027–1030
- Williams, P. and K. Norris (1987). *Near-Infrared Technology in the Agricultural and Food Industries*
- Wold, S., M. Sjöström, and L. Eriksson (2001). PLS-Regression: A Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **58**(2); 109–130
- You, H., B. Ireland, M. Moeszinger, H. Zhang, L. Snow, S. Krepich, and V. Takagawa (2019). Determination of Bioactive Nonvolatile Ginger Constituents in Dietary Supplements by a Rapid and Economic HPLC Method: Analytical Method Development and Single-Laboratory Validation. *Talanta*, **194**; 795–802
- Yu, D. X., S. Guo, X. Zhang, H. Yan, Z. Y. Zhang, X. Chen, J. Y. Chen, S. J. Jin, J. Yang, and J. A. Duan (2022). Rapid Detection of Adulteration in Powder of Ginger (*Zingiber officinale* Roscoe) by FT-NIR Spectroscopy Combined with Chemometrics. *Food Chemistry: X*, **15**
- Yücel, Ç., G. Ş. Karatoprak, Ö. B. Açıkar, E. K. Akkol, T. H. Barak, E. Sobarzo-Sánchez, M. Aschner, and S. Shirooie (2022). Immunomodulatory and Anti-Inflammatory Therapeutic Potential of Gingerols and Their Nanoformulations. *Frontiers in Pharmacology*, **13**; 902551
- Yudha S., S., E. Angasa, M. A. Reagen, and M. Kazi (2024). Molecular Spectroscopic (FTIR and UV-Vis) Analysis and *In Vitro* Antibacterial Investigation of a Deep Eutectic Solvent of N,N-Dimethyl Urea-Citric Acid. *Science and Technology Indonesia*, **9**(1); 167–172
- Zhang, S., X. Kou, H. Zhao, K. K. Mak, M. K. Balijepalli, and M. R. Pichika (2022). *Zingiber officinale* var. *rubrum*: Red Ginger's Medicinal Uses. *Molecules*, **27**(3); 775