JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 64-80, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

Research Article

OPEN ACCESS Check for updates

Evaluation of Catechin Content in Gambir Leaf Herbal Tea Using NIR Spectroscopy with PLS and MLR

Santos, Andasuryani*, Ifmalinda

Department of Agricultural Engineering and Biosystem, Faculty of Agricultural Technology, Andalas University, Limau Manis Campus, Padang 25163, West Sumatera, Indonesia.
*Corresponding author, email: andasuryani@ae.unand.ac.id

Article Info

Submitted: 22 December 2026
Revised: 20 February 2026
Accepted: 10 March 2026
Available online: 9 April 2026
Published: March 2026

Keywords:
calibration models, catechin, gambir leaves, NIRS

How to cite:
Santos., Andasuryani., Ifmalinda. (2026). Evaluation of Catechin Content in Gambir Leaf Herbal Tea Using NIR Spectroscopy with PLS and MLR. Jurnal Keteknikan Pertanian, 14(1): 64-80. https://doi.org/*10.19028/jtep.014.1.64-80.*

Abstract

Several factors, including drying duration, leaf maturity level, and drying temperature, can influence the catechin content of gambir herbal tea. Therefore, evaluating the quality of gambir herbal tea is essential. The commonly used method involves laboratory chemical analysis, which is destructive, generates chemical waste, and requires considerable time. The objective of this study was to develop an NIR model capable of rapidly, nondestructively, and environmentally friendly predicting the catechin content of gambir herbal tea. The reflectance of herbal tea samples was measured using a Büchi NIRFlex N500 spectrophotometer over 1000–2500 nm. The catechin content was determined through chemical analysis using ethanol as the solvent. Spectral data were calibrated against the actual catechin values using Partial Least Squares (PLS) and Multiple Linear Regression (MLR) in The Unscrambler X software. The pretreatments applied included Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), and Baseline correction. The best model was obtained using the MLR method combined with MSC pretreatment, yielding $Rc^2 = 0.91$, $Rp^2 = 0.92$, RMSEC = 1.78, RMSEP = 1.79, SEC = 1.69, SEP = 1.77, RPD = 3.44, and RER = 10.77. The PLS model with Baseline pretreatment also produced reliable predictions with $Rc^2 = 0.93$, $Rp^2 = 0.89$, RMSEC = 1.50, RMSEP = 2.09, SEC = 1.52, SEP = 2.08, RPD = 3.00, and RER = 9.56. This study successfully developed an NIR model capable of predicting catechin content in gambir leaves herbal tea with strong predictive performance.

Doi: https://doi.org/10.19028/jtep.014.1.64-80

1. Introduction

Gambir (Uncaria gambir Roxb.) is a plant from the Rubiaceae family that grows abundantly in Sumatra (Hayati et al., 2020). This plant is commonly used in traditional medicine for its antibacterial properties, as well as its antioxidant and antihyperlipidemic effects (Alegantina and Setyorini, 2017; Widiyarti et al., 2020). These therapeutic effects originate from various compounds contained in gambir, including catechins, tannins, quercetin, quinic acid, procyanidin dimers, and

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

proanthocyanidin dimers (Ismail et al., 2021). Gambir is typically produced by boiling its stems and leaves to extract the sap, which is then processed into powder and blocks (Yeni et al., 2014). These forms are utilized in the pharmaceutical, textile, and food industries. However, consumer preferences have shifted toward gambir tea, valued for its convenience and immediate enjoyment, with no need for complex preparation.

Gambir leaves are used to make herbal tea, which is made from processed, dried gambir leaves and contains catechin as its main compound, a polyphenol derivative classified as a type of condensed tannin (Tavita et al., 2023). Catechins offer numerous benefits, including anti-inflammatory, anti-carcinogenic, antioxidant, and antibacterial properties (Chen et al., 2016; Isemura, 2019). The quality of herbal tea is determined by its catechin content; higher levels indicate better quality and greater potential for success in international markets. Factors that influence catechin levels during the processing of gambir leaves include leaf maturity, drying time, and drying temperature. Therefore, evaluating the quality of gambir leaves herbal tea is essential to ensure that it meets consumer standards.

Evaluating catechin content in gambir leaves herbal tea typically involves chemical laboratory analyses, such as HPLC, spectrophotometry, ethyl acetate extraction, and ethanol extraction. However, these processes are inefficient, require considerable effort and time, and generate hazardous chemical waste from solvent-contaminated samples (Widyaningrum et al., 2022). Therefore, a faster and waste-free method for determining catechin content is needed.

Near-Infrared (NIR) spectroscopy uses electromagnetic waves with wavelengths ranging from 780 to 2500 nm to analyze hydrogen (H) chemical bonds with other elements (X), including carbon (C), oxygen (O), and nitrogen (N). This technology offers rapid analysis (Zhang et al., 2023) while generating minimal hazardous chemical waste (Zou et al., 2022). Currently, NIR spectroscopy is applied in the agricultural, pharmaceutical, petrochemical, and textile sectors (Wang et al., 2023). NIR spectra consist of absorbance and reflectance data that provide information about the chemical bonds of the analyzed elements; however, these signals often overlap with those from other molecular bonds. Therefore, spectral data processing requires multivariate analytical methods capable of extracting relevant bond information associated with the target compounds, such as Partial Least Squares (PLS) and Multiple Linear Regression (MLR).

Partial least squares (PLS) is a latent variable-based analytical method designed to model the relationship between a set of predictor or independent variables and a dependent variable. PLS is widely used in NIR spectral data processing because it can effectively handle multicollinearity and the high dimensionality of NIR datasets (Sousa and Åberg, 2018). The development of the NIR model for catechin content was previously conducted by Andasuryani et al. (2013), using PLS regression, and the resulting model had good predictive ability with a coefficient of determination (R^2) of 0.871. Zhang et al. (2021) employed the PLS method to develop a predictive model for active compounds in

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

green tea, achieving an R^2 of 0.98. Similarly, Dong et al. (2017) applied PLS to predict catechins (0.909), theaflavins (0.891), and polyphenols (0.93) in black tea. Multiple linear regression (MLR) is an analytical technique used to examine the relationship between a dependent variable and multiple independent variables. MLR offers a more straightforward interpretative approach by enabling the selection of specific wavelengths as predictor variables (Beć et al., 2025). This characteristic contributes to the parsimony or computational simplicity of MLR models (Plonsky, 2017). The application of MLR to NIR spectral analysis was demonstrated by Firdaus et al. (2024), who used it to predict macronutrient content in paddy soils. Their results showed that the model successfully predicted nitrogen (N), phosphorus (P), and potassium (K) levels, with R^2 values of 0.86, 0.85, and 0.94, respectively. However, no studies have yet explored the use of NIR spectroscopy to predict catechin content in gambir leaves herbal tea. Therefore, further research using NIR calibration models is necessary.

2. Material and Methods

2.1 Sample Preparation

The gambir leaves used in this study were obtained from the Cubadak variety and sourced from Siguntur Village, Pesisir Selatan Regency, West Sumatra. After being washed with clean water, the leaves were withered at room temperature for approximately 17 h (Anggraini, 2017). They were then dried in a rack-type oven at 45 °C (Dorkbuakaew et al., 2016) and chopped using a chopper. The resulting powder was sieved using a 40-mesh screen. A total of 60 samples were prepared, each weighing ± 22 grams. 40 samples were allocated for develop the calibration model, while the remaining 20 samples were used to validate it. The sample division method followed the Kennard–Stone sampling algorithm (Ren et al., 2021).

2.2 Spectral Acquisition

A powdered sample weighing ± 22 g was placed in a Petri dish for spectral data collection. Each sample was scanned three times using near-infrared (NIR) light, over 1000–2500 nm (10000–4000 cm-1) at 4 cm-1 intervals, and the resulting spectra were averaged. The acquired spectral data were recorded as reflectance and subsequently transformed into absorbance using Log(1/R). The NIR instrument was operated using NIR Ware 1.2 software. The spectrum acquisition process is shown in Figure 1.

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 64-80, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

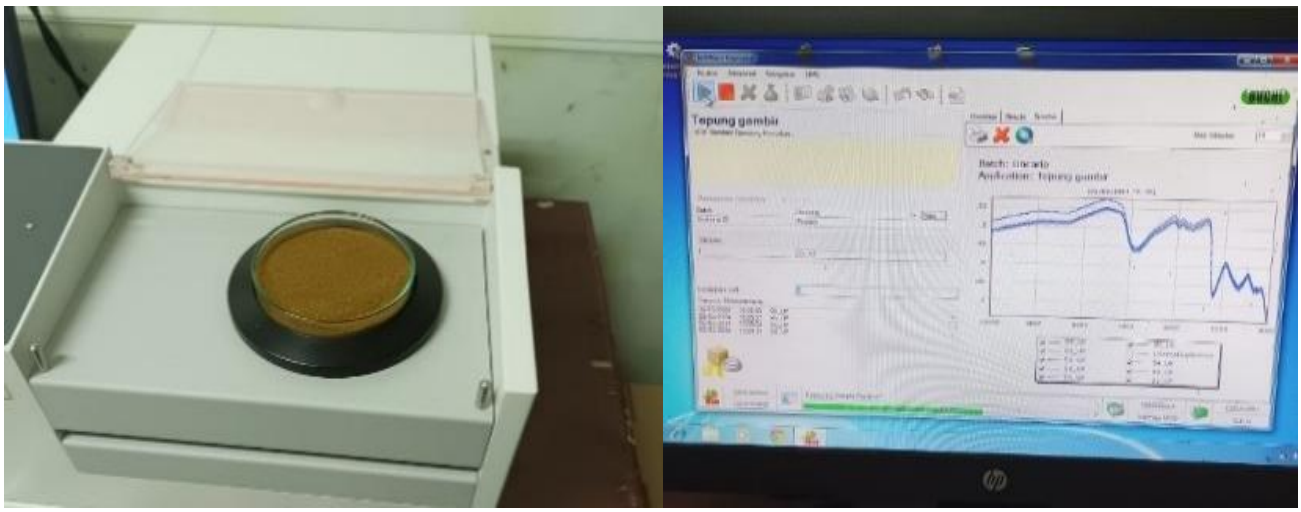

Figure 1. Spectral Acquisition of Herbal Tea Gambir Leaf.

2.3 Chemical Analysis

The catechin content was determined through laboratory chemical analysis. A herbal gambir leaf tea extract solution was prepared at a concentration of 1000 μg/mL to obtain a total volume of 100 mL. Subsequently, 1 mL of this solution was pipetted and diluted to 10 mL. Subsequently, 0.5 mL of the diluted solution was transferred into a 10 mL volumetric flask and adjusted to the desired volume. The absorbance of the sample was measured at 279 nm using a UV-visible spectrophotometer (Nur et al., 2020).

2.4 Chemometrics Methods

The data were analyzed using Unscrambler X software to develop calibration models. Before model development, spectral pretreatments were applied to reduce noise and enhance spectral quality. Calibration models were constructed using two multivariate calibration methods: Partial Least Squares (PLS) and Multiple Linear Regression (MLR). In PLS, predictor variables were determined by reducing the spectral data into several latent variables, which were then incorporated into regression equations (Amirul et al., 2020). The correct number of latent variables is indicated by a small RMSE and a high coefficient of determination in calibration and internal validation (Chariskou et al., 2022). In contrast, the MLR method selects predictor variables by identifying specific wavelengths to be included in the model. Several wavelength selection techniques can be used; however, this study employed the stepwise method. The advantage of stepwise selection is that all variables are sequentially tested in the regression equation to evaluate their significance (Hahs-Vaughn, 2016).

The statistical parameters used to evaluate the models included the coefficient of determination for calibration (Rc^2), the coefficient of determination for prediction (Rp^2), standard error of calibration (SEC), standard error of prediction (SEP), root mean squared error of calibration (RMSEC), root mean

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 64-80, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

squared error of prediction (RMSEP), ratio of performance to deviation (RPD), and Range Error Ratio (RER). A model is considered highly precise when it achieves an R^2 value greater than 0.9, whereas models with R^2 values below 0.29 are regarded as poor and not recommended for use (Martín, 2022). According to Heil and Schmidhalter (2021), a good model should have an RPD value greater than 3; an RPD value below 3 indicates that the model performs only coarse predictions. Models with RPD ≤ 1.9 are considered poor predictors and are not recommended for application. The RMSE value suggests the accuracy of the model's calculations, whereas SEP and SEC represent the precision of the prediction and calibration processes. RER is the ratio between the range of reference data and error parameters, such as RMSEC or RMSEP (Yang et al., 2017). The equations for calculating R^2, RMSE, RPD, and RER are presented in Equations 1, 2, 3, and 4, respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(\bar{y}_i - \hat{y}_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}i - yi)^2}{n}} \quad (2)$$

$$RPD = \frac{SD}{RMSE_{prediction}} \quad (3)$$

$$RER = \frac{ymax - ymin}{RMSE_{prediction}} \quad (4)$$

Where: y, $\hat{y}$, and $\bar{y}$ is the reference, predicted, and average of catechin content, respectively, n is the number of samples for calibration or prediction, and SD is the standard deviation.

3. Results and Discussion

3.1 Spectra of Gambir Leaves Herbal Tea

The NIR spectra of herbal gambir leaf tea were measured using an NIR instrument within a wavelength range of 1000–2500 nm. Figure 2 presents the raw spectral data, which show noticeable gaps between spectra due to the absence of pretreatment. The NIR spectra of herbal gambir leaf tea exhibit characteristic absorption patterns of polyphenolic compounds, particularly catechins, which contain multiple aromatic hydroxyl groups. A strong absorption peak appears at approximately 1450 nm, corresponding to the first overtone of O–H vibrational stretching. A prominent peak observed at 1900–1930 nm corresponds to the O–H combination band, which is typically highly sensitive to the moisture content of the sample. In catechin-rich matrices, this band becomes dominant because of hydrogen bonding between phenolic O–H groups and water, thereby increasing the absorption intensity. This observation is supported by Mardiantono et al. (2022), who reported that the absorption peak corresponding to O–H bonding, commonly identified as the primary water absorption peak, occurs at approximately 1450 nm. This is in accordance with the absorption features observed in the

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

1400–1500 nm range. Additionally, combination absorption bands are clearly visible within the 1900–1950 nm range, which is associated with chemical constituents in herbal tea, such as C–O and O–H bonds, which effectively absorb NIR light.

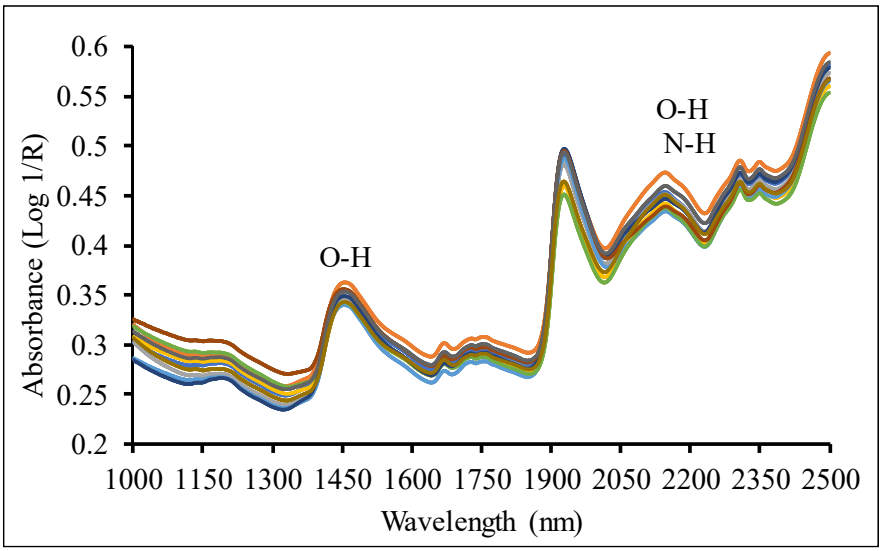

Figure 2. Original NIR Spectra of Gambir Leaf Herbal Tea.

Workman and Weyer (2007) reported that the absorption observed at 1700 nm originates from the overtone of aromatic and aliphatic C–H bonds attached to the flavan-3-ol backbone. The bands in the 2140–2260 nm region can be attributed to the combination vibrations of C–H and phenolic C–O groups. The presence of these bands indicates contributions from the aromatic structure of catechins, which contain abundant C–O and C–H bonds. At higher wavelengths, absorption in the 2440–2500 nm range reflects the strong combination bands of O–H and C–H, commonly observed in organic compounds with phenolic functional groups. Overall, the spectral features demonstrate that catechin, as the primary component of gambir leaves, significantly contributes to the NIR absorption pattern.

Figure 3 shows the regression coefficients of each wavelength in the spectra. Wavelengths from 1650 nm to 1700 nm contribute the most to the catechin content. Furthermore, at wavelengths of 1713.5 nm, 1948.56 nm, 1976.28 nm, and 2144.08 nm, the catechin composition is moderate. Finally, wavelengths with low catechin content start from 1753.16 nm, 2050.86 nm, 2252.25 nm, 2304.15 nm, 2351.83 nm, and 2399.23 nm. Overall, all wavelengths influence the spectral intensity. Almost all of these wavelengths are absorption peaks, although the absorption intensities are low.

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 64-80, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

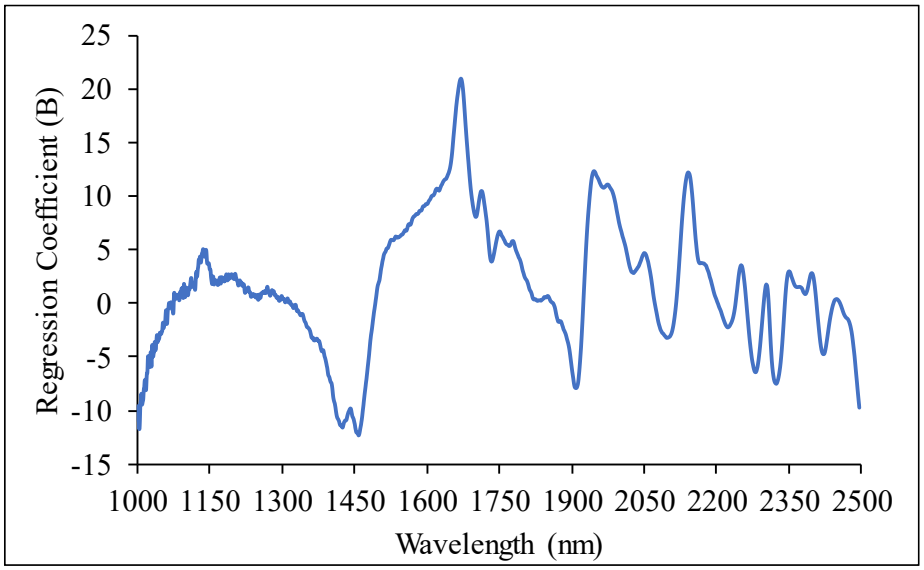

Figure 3. Regression Coefficient of Each Wavelength in the Spectra.

3.2 Regression Coefficient of Each Wavelength in the Spectra

The application of pretreatments to NIR spectral data aims to reduce non-chemical effects, such as light scattering caused by particle size, variations in sample density, and instrumental baseline shifts. In the raw spectra, noticeable baseline differences are observed among samples, particularly in the higher-wavelength region (>2000 nm). Such variations can reduce model accuracy because chemical information becomes mixed with scattering noise. Workman and Weyer (2007) noted that baseline shifts and scattering effects resulting from particle size and sample packing are among the most common sources of variation in raw NIR spectra. The results of the SNV pretreatment are shown in Figure 4. The graph shows substantial changes in absorbance, with both positive and negative values, indicating that the spectra have been normalized relative to their means. According to Budiastra et al. (2020), SNV is highly effective at eliminating disturbances caused by scattering and particle-size effects, thereby making sample spectra easier to distinguish. This enhancement makes the main peaks (1450 nm and 1900 nm) more representative of the chemical content, thereby improving the accuracy of the calibration models.

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

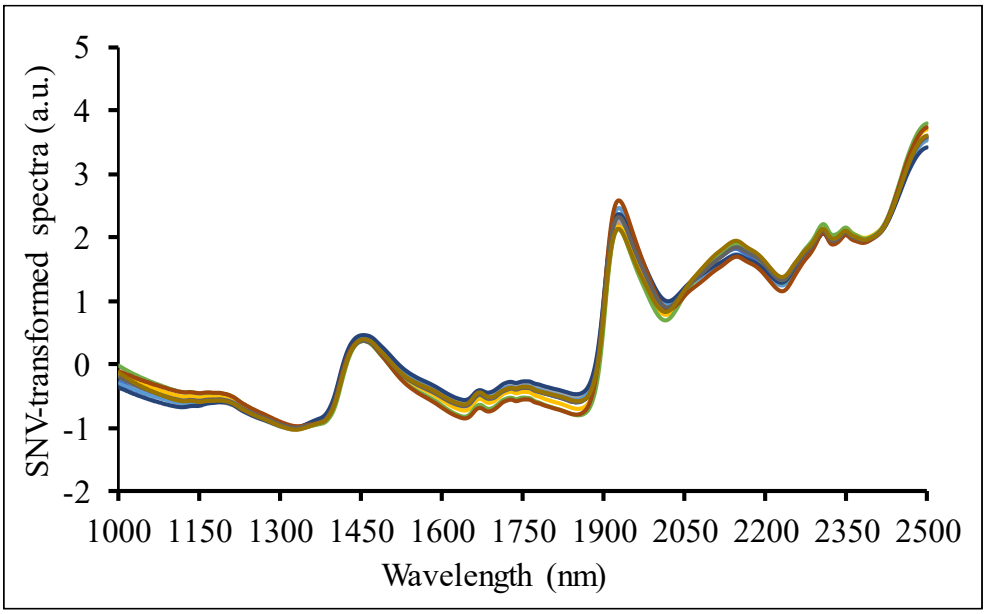

Figure 4. Gambir Leaves Herbal Tea Spectra with SNV Pretreatment.

Figure 5 shows the spectra corrected using baseline pretreatment. The spectral profiles become flatter at lower absorbances, indicating a reduced baseline drift. This approach primarily corrects spectral shifts caused by instrumental variations. As noted by Rinnan et al. (2009), baseline correction is applied to reduce additive effects and instrument-related reading errors; however, it is not effective in eliminating scattering effects. Figure 6 displays the results of the MSC pretreatment, which produces spectra that closely resemble the raw spectra but with a more uniform baseline across samples. This technique works by adjusting the slope and offset of each spectrum relative to the mean spectrum. This is in accordance with the statement by Workman and Weyer (2007) that MSC modifies each spectrum based on a reference by correcting its slope and offset, thereby reducing scattering effects while preserving the original spectral characteristics. Consequently, MSC can enhance the robustness of calibration models without altering the main chemical absorption features. Overall, SNV provides the strongest correction for scattering variations, making it particularly effective for powdered samples, such as gambir leaves. MSC maintains the natural spectral shape while reducing scatter effects, whereas baseline correction is simpler and particularly useful for minimizing instrumental drift.

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 64-80, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

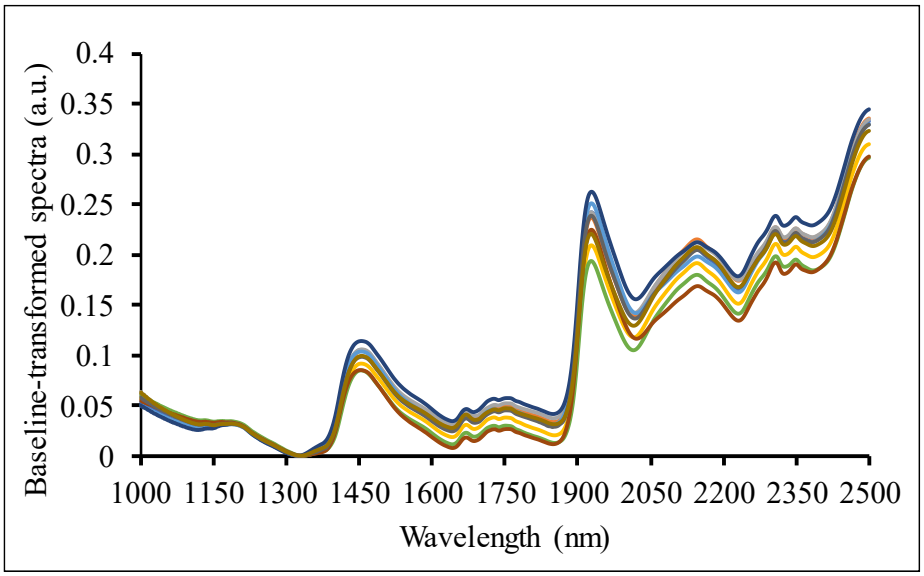

Figure 5. Gambir Leaves Herbal Tea Spectra with Baseline Pretreatment.

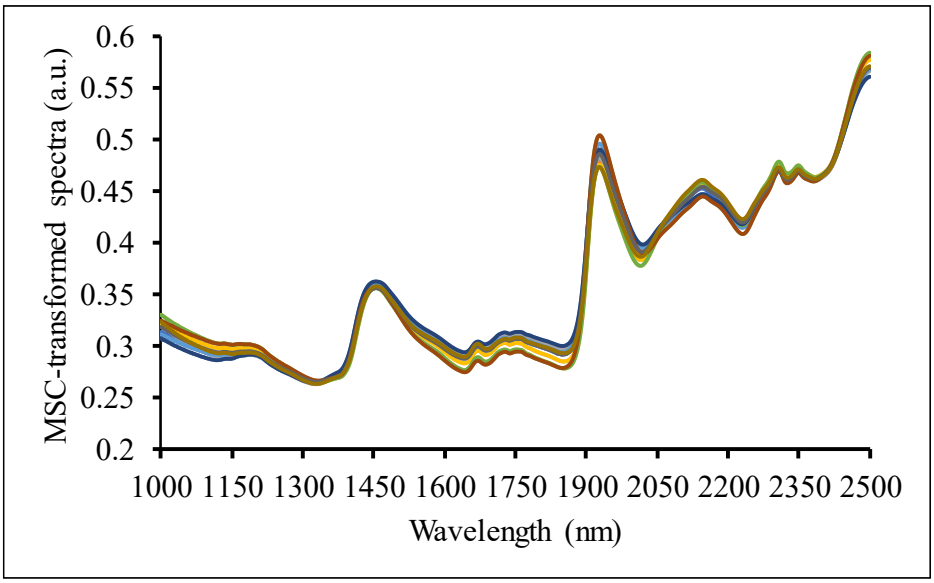

Figure 6. Gambir Leaves Herbal Tea Spectra with MSC Pretreatment.

3.3 Calibration and Validation Results of the Model Using Multivariate Analysis PLS dan MLR

The statistical results of the model calibration and validation are presented in Table 1. Based on calibration using the Partial Least Squares (PLS) method, the original spectra showed reasonably good performance, with R^2c and R^2p values of 0.95 and 0.86, respectively. However, the RPD value of 2.65 indicates that the model still falls within the acceptable category for coarse or preliminary prediction (Heil and Schmidhalter, 2021). After applying various pretreatments, the model performance showed noticeable variations. The Standard Normal Variate (SNV) pretreatment slightly reduced the model's

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 64-80, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

performance, as reflected by a decrease in R^2c to 0.92 and R^2p to 0.83, along with an RPD value of 2.32. This decline suggests that pretreatments do not always guarantee an improvement in model quality. This result aligns with that of Cen and He (2007), who noted that SNV can enhance spectral differences but may also reduce calibration performance if some important chemical information in the spectra is inadvertently removed.

Table 1. Statistical Results of Calibration and Validation Models Using Various Pretreatment Methods.

Pretreatment	Number of Factors	Calibration			Validation			RPD	RER
		R^2c	RMSEC (%)	SEC (%)	R^2p	RMSEP (%)	SEP (%)		
PLS									
Original	7	0.95	1.23	1.24	0.86	2.48	2.48	2.65	7.77
SNV	5	0.92	1.56	1.58	0.83	2.62	2.67	2.32	7.36
Baseline*	**6**	**0.93**	**1.5**	**1.52**	**0.89**	**2.09**	**2.08**	**3**	**9.56**
MSC	6	0.93	1.43	1.45	0.88	2.16	2.11	2.82	8.93
MLR									
Original	4	0.89	1.97	1.87	0.89	1.95	1.98	3.1	9.89
SNV	5	0.9	1.94	1.81	0.91	1.82	1.81	3.32	10.59
Baseline	5	0.95	1.32	1.23	0.89	1.99	2.04	3.12	10.10
MSC**	**4**	**0.91**	**1.78**	**1.69**	**0.92**	**1.79**	**1.77**	**3.4**	**10.77**

* The best performance model from PLS method

** The best performance model from both methods

The best-performing model was obtained using the baseline pretreatment method, yielding R^2c = 0.93 and R^2p = 0.89, indicating high predictive precision. In addition, the SEC and SEP values obtained were not significantly different from those produced by other pretreatment methods. The RER value achieved using the baseline pretreatment was 9.56, the highest among all tested methods. This value falls within the range of 3–10, indicating a model with limited capability. According to Quentin et al. (2017), models with RER values between 3 and 10 have limited applicability, whereas models with RER > 10 are considered highly useful. Meanwhile, MSC produced balanced R^2c and R^2p values of 0.93 and 0.88, respectively, with an RPD value of 2.82. This demonstrates that MSC can preserve chemical information while reducing scattering effects. Workman and Weyer (2007) explained that MSC corrects the slope and offset of spectra, reducing scattering while maintaining the original spectral characteristics.

Andasuryani et al. (2014) also developed an NIR model using PLS regression to predict catechin content in gambir powder and gambir blocks. The best-performing models used seven latent variables for gambir powder and six latent variables for gambir blocks. The resulting R^2 values were 0.83 and

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

0.91 for gambir powder and gambir blocks, respectively. In addition, the SEP values obtained for both forms were relatively similar, indicating comparable prediction errors. The SEP value should not exceed twice the SEC value; if this condition is violated, the model may be considered overfitted. In this study, the SEP values obtained from models using all pretreatment methods were generally greater than twice but did not exceed twice the corresponding SEC values. Therefore, the PLS models can be considered to have good overall performance without indications of overfitting. The RMSEC values for each pretreatment method also varied but remained relatively low compared with the RMSEP values. The RMSEC values, ranging from 1.23% to 1.56%, indicate good model accuracy with minimal prediction errors in the calibration dataset. PLS is more sensitive to noise in the spectra and may exhibit mild overfitting when the number of latent factors is not appropriately selected (Reyna et al., 2017; Sousa and Åberg, 2018).

The MLR method produced models with better and more varied performance than the PLS method. The raw spectra used wavelengths of 1914.24 nm, 1965.41 nm, 2069.54 nm, and 2136.75 nm, generating a reasonably strong model, with R^2c of 0.89 and R^2p of 0.89, and an RPD value of 3.1, which classifies the model as having good predictive ability. However, the RER value fell within the 3–10 range, indicating limited applicability of the model (Murguzur et al., 2021). The best model was obtained with the MSC pretreatment, achieving R^2c and R^2p of 0.91 and of 0.92, respectively, placing it in the highly precise category. The specific wavelengths used by the MSC method start from 1914.24 nm, 1953.13 nm, 2149.61 nm, and 2394.64 nm. The RPD value of 3.4 was the highest among all pretreatments, indicating a good predictive model (RPD > 3). Additionally, the RER value exceeded 10, indicating that the model is highly suitable for screening purposes (Crespo et al., 2024). This confirms that MSC enhances the reliability of simple linear models by reducing non-chemical variations. Conversely, although the baseline method yielded a high R^2c value of 0.95, its validation performance decreased, with an R^2p of 0.89. The baseline used wavelengths of 1057.08 nm, 1076.19 nm, 1442.59 nm, 1463.7 nm, and 1670.01 nm. The RPD value of 3.12 was also slightly lower than that of the MSC method. Meanwhile, the SNV method used wavelengths from 1361.66 nm, 1912.78 nm, 1956.18 nm, 2106.15 nm, and 2149.61 nm, achieved strong model performance, with a R^2c and R^2p of 0.90 and 0.91, respectively. Its RPD value was comparable to that of MSC, though slightly lower. Overall, these findings indicate that pretreatment significantly affects model quality. The baseline correction was most effective for PLS, increasing the RPD to 3.0, whereas MSC performed best for MLR with an RPD of 3.4. This is consistent with the findings of Zhang et al. (2022), who emphasized that no single pretreatment method is universally optimal; rather, its selection should be tailored to the data characteristics and calibration method used. The relationship between reference catechin values and predicted catechin values using the PLS method is shown in Figure 7.

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 64-80, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

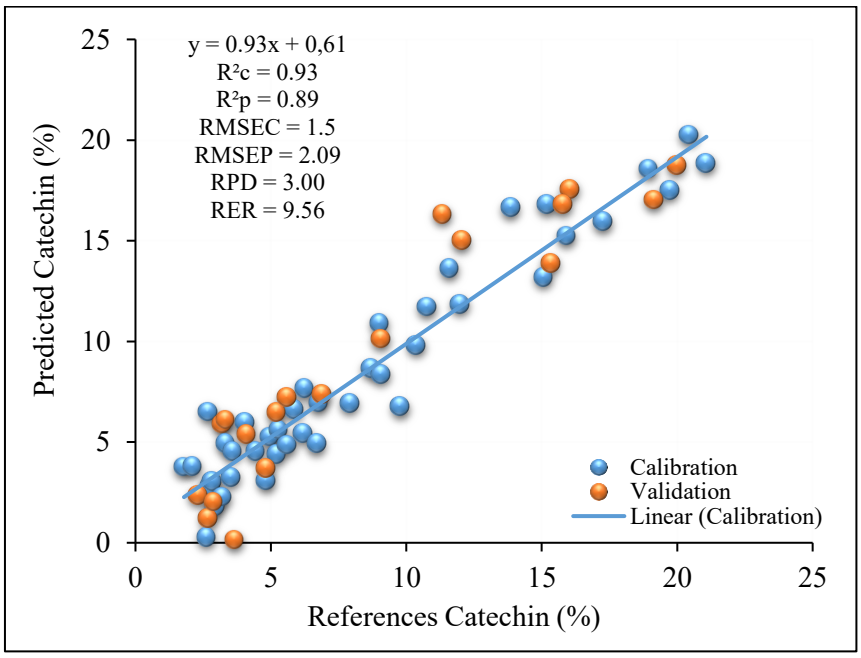

Figure 7. Plot of the Relationship Between NIR Predictions and Reference Data Using the PLS Method.

Figure 8 shows the relationship between the reference and predicted catechin values obtained using the MLR method. The MLR model combined with MSC is more stable and exhibits stronger generalization capability when applied to new data. The estimated RMSEC and RMSEP values are relatively close, as are the SEC and SEP values, indicating that the prediction errors are well-balanced between calibration and validation. PLS improves the accuracy of the training data but is more prone to overfitting. This is in accordance with the known characteristics of PLS, which is more sensitive to noise in the spectra and may experience mild overfitting when the number of latent factors is not selected appropriately (Reyna et al., 2017; Sousa and Åberg, 2018).

Meanwhile, MLR combined with MSC tends to produce more stable predictions, although its calibration accuracy is slightly lower. Therefore, for practical applications such as the nondestructive determination of catechin quality in gambir leaves, MLR with MSC is recommended. At the same time, PLS remains relevant for exploratory studies that prioritize high calibration accuracy.

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 64-80, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

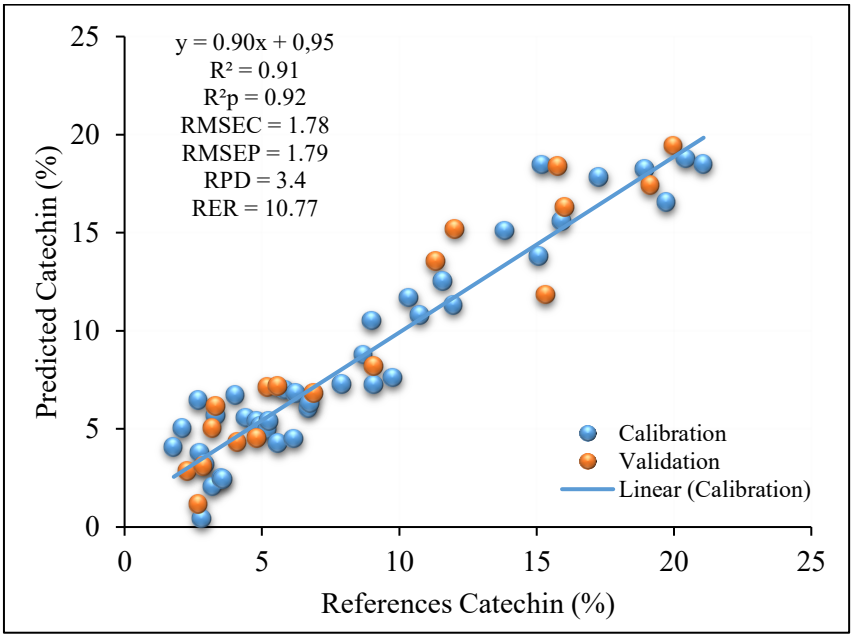

Figure 8. Plot of the Relationship Between NIR Predictions and Reference Data Using the MLR Method.

4. Conclusion

The results of this study demonstrate that NIR spectroscopy combined with chemometric methods can nondestructively predict catechin content in gambir leaves with high accuracy. The PLS model with baseline pretreatment produced the highest coefficient of determination (R^2) during calibration, reaching 0.93. However, its performance declined during validation, with an R^2p of 0.89 and a larger error gap, indicating a slight risk of overfitting. In contrast, the multiple linear regression (MLR) model with MSC pretreatment showed more stable performance, with an R^2c of 0.91 during calibration and an R^2p of 0.92 during validation, along with more consistent error values. Therefore, although PLS provides the highest calibration accuracy, the MLR model with MSC pretreatment is more suitable for practical applications in evaluating catechin quality in gambir leaves, as it delivers more stable and reliable predictions during validation.

5. Acknowledgement

The authors express their sincere gratitude to Prof. Dr. Andasuryani, S.TP., M.Si. for funding this research. Appreciation is also extended to Raisal, Alamsyah, Regi, and Fuja for their valuable assistance and support throughout the study.

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

6. AI Writing Statement

The author used the Generative AI Tool (ChatGPT 4.5) to translate the manuscript. The analysis, discussion, and conclusions are the result of the author's thinking.

7. References

Alegantina, S., and Setyorini, H. A. (2017). Gambaran Cemaran dan Kadar Metil Galat pada Tiga Mutu Ekstrak Gambir (Uncaria gambir Roxb.). Jurnal Kefarmasian Indonesia, 7(1), 10–18. https://doi.org/10.22435/jki.v7i1.5876.10-18

Amirul, M. S., Endut, R., Rashidi, C. B. M., Aljunid, S. A., Ali, N., Laili, M. H., Laili, A. R., and Ismail, M. N. M. (2020). Estimation of Harumanis (Mangifera indica L.) Sweetness using Near-Infrared (NIR) Spectroscopy. IOP Conference Series: Materials Science and Engineering, 767(1). https://doi.org/10.1088/1757-899X/767/1/012070

Andasuryani, Purwanto, Y. A., Budiastra, I. W., and Syamsu, K. (2014). Prediksi Kandungan Katekin Gambir (Uncaria gambir Roxb.) dengan Spektroskopi NIR. J Tek Ind Pert, 24(1), 43–52. http://insightsociety.org/ojaseit/index.php/ijaseit/article/view/423/465

Andasuryani, Syamsu, K., Purwanto, Y. A., and Budiastra, I. W. (2013). Application of near-infrared reflectance spectroscopy for determination of Gambir (Uncaria gambir Roxb.) Moisture Andcatechin content. Acta Horticulturae, 1011, 299–306.

Anggraini, T. (2017). Proses dan Manfaat Teh. erka.

Beć, K. B., Grabska, J., and Huck, C. W. (2025). Interpretability in Near-Infrared (NIR) Spectroscopy: Current Pathways to the Long-Standing Challenge. TrAC - Trends in Analytical Chemistry, 189(118254), 1–17. https://doi.org/https://doi.org/10.1016/j.trac.2025.118254

Budiastra, I. W., Sutrisno, M., Widyotomo, S., and Ayu, P. C. (2020). Determination of Trigonelline and Chlorogenic Acid (CGA) Concentration in Intact Coffee Beans by NIR Spectroscopy. Agricultural Engineering International : The CIGR e-Journal, 22(1), 162–168. https://cigrjournal.org/index.php/Ejounral/article/view/4590

Cen, H., and He, Y. (2007). Theory and Application of Near Infrared Reflectance Spectroscopy in Determination of Food Quality. Trends in Food Science and Technology, 18(2), 72–83. https://doi.org/doi:10.1016/j.tifs.2006.09.003

Chariskou, C., Vrochidou, E., and Daniels, A. J. (2022). Variable Selection on Reflectance NIR Spectra for the Prediction of TSS in Intact Berries of Thompson Seedless Grapes. Agronomy, 12(2113), 1–11. https://doi.org/doi.org/10.3390/agronomy12092113

Chen, X. Q., Hu, T., Han, Y., Huang, W., Yuan, H. B., Zhang, Y. T., Du, Y., and Jiang, Y. W. (2016). Preventive effects of catechins on cardiovascular disease. Molecules, 21(12), 1–7. https://doi.org/10.3390/molecules21121759

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

Crespo, S. P., Gesteiro, N., López-Malvar, A., Gómez, L., and Santiago, R. (2024). Assessing the Application of Near-Infrared Spectroscopy to Determine Saccharification Efficiency of Corn Biomass. Bioenergy Research, 17(3), 1522–1532. https://doi.org/doi.org/10.1007/s12155-024-10761-4

Dong, C., Zhu, H., Wang, J., Yuan, H., Zhao, J., and Chen, Q. (2017). Prediction of black tea fermentation quality indices using NIRS and nonlinear tools. Food Science and Biotechnology, 26(4), 853–860. https://doi.org/10.1007/s10068-017-0119-x

Dorkbuakaew, N., Ruengnet, P., Pradmeeteekul, P., Nimkamnerd, J., Nantitanon, W., and Thitipramote, N. (2016). Bioactive Compounds and Antioxidant Activities of Camellia sinensis var. assamica in Different Leave Maturity from Northern Thailand. In International Food Research Journal (Vol. 23, Issue 5, pp. 2291–2295). https://api.semanticscholar.org/CorpusID:201930007

Firdaus, J., Ahmad, U., Budiastra, I. W., and Subrata, I. D. M. (2024). Estimating Macronutrient Content of Paddy Soil Based on Near-Infrared Spectroscopy Technology Using Multiple Linear Regression. Advances in Technology Innovation, 9(1), 50–64. https://doi.org/https://doi.org/10.46604/aiti.2023.12683

Hayati, P. K. D., Fitri, A. W., and Fauza, H. (2020). Flowering characterization and its relation to the type of pollination on gambier (Uncaria gambir (Hunter) Roxb.). IOP Conference Series: Earth and Environmental Science, 583(1). https://doi.org/10.1088/1755-1315/583/1/012001

Heil, K., and Schmidhalter, U. (2021). An evaluation of different nir-spectral pre-treatments to derive the soil parameters c and n of a humus-clay-rich soil. Sensors, 21(1423). https://doi.org/10.3390/s21041423

Isemura, M. (2019). Catechin in human health and disease. Molecules, 24(3). https://doi.org/10.3390/molecules24030528

Ismail, A. S., Rizal, Y., Armenia, and Kasim, A. (2021). Identification of bioactive compounds in gambier (Uncaria gambir) liquid by-product in west Sumatra, Indonesia. Biodiversitas, 22(3), 1474–1480. https://doi.org/10.13057/BIODIV/D220351

Mardiantono, Budiastra, I. W., and Sutrisno. (2022). Prediksi Kandungan Kimia Kopra dengan FT-NIR Spectroscopy Menggunakan PLS. JTEP, 10(2), 87–97. https://doi.org/doi.org/10.19028/jtep.10.2.87-97

Murguzur, F. J. A., Tarroux, A., Bråthen, K. A., Bustamante, P., and Descamps, S. (2021). Using Near-Infrared Reflectance Spectroscopy (NIRS) to Estimate Carbon and Nitrogen Stable Isotope Composition in Animal Tissues. Ecology and Evolution, 11(15), 10483–10488. https://doi.org/10.1002/ece3.7851

Nur, S., Rumpak, G., Mubarak, F., Megawati, Aisyah, A. N., Marwati, Sami, F. J., and Fatmawaty, A. (2020). Identifikasi dan Penentuan Kadar Katekin dari Seduhan dan Ekstrak Etanol Produk Teh Hijau (Camelia sinensis L.) Komersial secara Spektofotometri UV-Visible. Majalah Farmasi Dan Farmakologi, 24(1), 1–4. https://doi.org/10.20956/mff.v24i1.9261

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.64-80

Plonsky, L. (2017). Advancing Quantitative Methods in Second (L. Plonsky (ed.)). Routledge.

Quentin, A. G., Rodemann, T., Doutreleau, M. F., Moreau, M., and Davies, N. W. (2017). Application of Near-Infrared Spectroscopy for Estimation of Non-Structural Carbohydrates in Foliar Samples of Eucalyptus globulus Labilladière. Tree Physiology, 37(1), 131–141. https://doi.org/10.1093/treephys/tpw083 Research

Ren, G., Ning, J., and Zhang, Z. (2021). Multi-Variable Selection Strategy Based on Near-Infrared Spectra for the Rapid Description of Dianhong Black Tea Quality. Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 245(118918), 1–9. https://doi.org/doi.org/10.1016/j.saa.2020.118918

Reyna, L., Dube, F., Barrera, J. A., and Zagal, E. (2017). Potential model overfitting in predicting soil carbon content by visible and near-infrared spectroscopy. Applied Sciences (Switzerland), 7(7). https://doi.org/doi:10.3390/app7070708

Rinnan, Å., Berg, F. van den, and Engelsen, S. B. (2009). Review of the Most Common Pre-processing Techniques for Near-Infrared Spectra. TrAC - Trends in Analytical Chemistry, 28(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Sousa, P. F. M., and Åberg, K. M. (2018). Can we beat overfitting?—A closer look at Cloarec's PLS algorithm. Journal of Chemometrics, 32(6), 1–10. https://doi.org/10.1002/cem.3002

Tavita, G. E., Sofiana, M. S. J., Ashari, A. M., Apindiati, R. K., Hartanti, L., and Warsidah, W. (2023). Characterization and Antioxidant Activity of Herbal Tea from Gambir Leaves (Uncaria gambir) with Different Drying Processes. Sainstek : Jurnal Sains Dan Teknologi, 15(2), 69. https://doi.org/https://doi.org/10.31958/js.v15i2.7719

Wang, S., Feng, L., Liu, P., Gui, A., Teng, J., Ye, F., Wang, X., Xue, J., Gao, S., and Zheng, P. (2023). Digital Prediction of the Purchase Price of Fresh Tea Leaves of Enshi Yulu Based on Near-Infrared Spectroscopy Combined with Multivariate Analysis. Foods, 12(19), 3592. https://doi.org/https://doi.org/10.3390/ foods12193592

Widiyarti, G., Sundowo, A., Filailla, E., and Arya Laksmono, J. (2020). The Mechanically Extraction Process of Gambier (Uncaria gambier Roxb.) from Limapuluh Kota, West Sumatera and Its Antioxidant activity. The Journal of Pure and Applied Chemistry Research, 9(1), 8–15. https://doi.org/10.21776/ub.jpacr.2020.009.01.509

Widyaningrum, W., Purwanto, Y. A., Widodo, S., Supijatno, and Iriani, E. S. (2022). Portable/Handheld NIR sebagai Teknologi Evaluasi Mutu Bahan Pertanian secara Non-Destruktif. Jurnal Keteknikan Pertanian, 10(1), 59–68. https://doi.org/10.19028/jtep.10.1.59-6

Workman, J., and Weyer, L. (2007). Practical Guide to Interpretive Near-Infrared Spectroscopy. In Practical Guide to Interpretive Near-Infrared Spectroscopy. CRC Press.

Yeni, G., Syamsu, K., Suparno, O., Mardliyati, E., and Muchtar, H. (2014). Repeated extraction process of raw gambiers (Uncaria gambier Robx.) for the catechin production as an antioxidant.

International Journal of Applied Engineering Research, 9(24), 24565–24578. http://www.ripublication.com

Zhang, S., Zuo, Y., Wu, Q., Wang, J., Ban, L., Yang, H., and Bai, Z. (2021). Development and Validation of Near-Infrared Methods for the Quantitation of Caffeine, Epigallocatechin-3-gallate, and Moisture in Green Tea Production. Journal of Analytical Methods in Chemistry, 2021, 1–12. https://doi.org/https://doi.org/10.1155/2021/9563162

Zhang, W., Kasun, L. C., Wang, Q. J., Zheng, Y., and Lin, Z. (2022). A Review of Machine Learning for Near-Infrared Spectroscopy. Sensors, 22(24), 1–32.

Zhang, Y., Wang, J., Luo, H., Yang, J., Wu, X., Wu, Q., and Zhong, Y. (2023). Rapid Prediction of Yongchuan Xiuya Tea Quality by Using Near Infrared Spectroscopy Coupled with Chemometric Methods. Food Science and Technology (Brazil), 43. https://doi.org/https://doi.org/10.1590/fst.101122 Rapid

Zou, H., Shen, S., Lan, T., Sheng, X., Zan, J., Jiang, Y., Du, Q., and Yuan, H. (2022). Prediction Method of the Moisture Content of Black Tea during Processing Based on the Miniaturized Near-Infrared Spectrometer. Horticulturae, 8(12). https://doi.org/https:// doi.org/10.3390/horticulturae8121170.