Rapid and Non-Destructive Prediction of Total Phenolic Content in Coffee-Leaf Powder Using Near-Infrared Spectroscopy and Chemometrics

Nadrah Puspita[ID], Andasuryani*[ID], Dinah Cherie[ID]

Agricultural Engineering Department, Andalas University, Padang 25163, Indonesia

Corresponding Author Email: andasuryani@ae.unand.ac.id

ABSTRACT

Rapid and reliable quality evaluation of coffee-leaf powder, particularly total phenolic content (TPC), is essential for large-scale screening and product standardization. Although conventional wet-chemistry methods are accurate, they are time-consuming, labor-intensive, and reagent-dependent, making them less practical for routine analysis of large sample sets. This study evaluated near-infrared (NIR) spectroscopy combined with chemometrics for non-destructive prediction of TPC in coffee-leaf powder. Ninety Robusta coffee-leaf powder samples were collected from several regions in West Sumatra, Indonesia. Reference TPC was determined using the Folin–Ciocalteu method, and reflectance spectra were acquired from 1000 to 2500 nm. Partial least squares regression (PLSR) models were developed using standard normal variate (SNV), multiplicative scatter correction (MSC), and their combinations with a second-order gap-segment derivative. The best-performing model (MSC + second-order derivative) used four latent variables and achieved $R^2_{cal} = 0.98$ and $R^2_{pred} = 0.97$, with residual predictive deviation (RPD) = 4.71 and range error ratio (RER) = 14.63 under external validation. These findings indicate that combining MSC with second-order derivative preprocessing likely improved model robustness and enhanced predictive performance for phenolic-related information. Overall, NIR spectroscopy combined with chemometrics shows strong potential as a rapid, non-destructive, and eco-friendly method for TPC screening in coffee-leaf powder.

1. INTRODUCTION

The coffee plant (*Coffea* sp.) has long been recognized for its economic significance, primarily through the utilization of its beans as the principal raw material for coffee beverages. More recently, sustainability-oriented approaches in the coffee sector have also emphasized the importance of diversifying coffee-derived products to enhance added value and resource utilization [1]. Against this background, coffee leaves have gained attention as a potential raw material for herbal infusion products. This potential is supported by their long-standing traditional use in several coffee-producing countries, as well as by reports of their antioxidant and anti-inflammatory activities [2]. In West Sumatra, Indonesia, this utilization has developed into a distinctive local beverage known as *kahwa* or *kawa daun*, in which the leaves are commonly smoked before being brewed in water [3, 4]. Furthermore, coffee leaves have been reported to exhibit sensory properties comparable to black tea and to be associated with relaxing properties, indicating their further promise as a functional drink [5].

Coffee leaves are known to contain various bioactive compounds, including polyphenols, saponins, mangiferin alkaloids, and flavonoids [6]. Among these constituents, polyphenols have received particular attention due to their role as natural antioxidants that are important for human health [7].

As the utilization of coffee leaves as herbal tea continues to increase, the need to quantify total phenolic content (TPC) becomes increasingly relevant, both for quality characterization and for comparisons across raw material conditions and processing treatments. Numerous studies indicate that the polyphenol content of coffee leaves can vary substantially. Acidri et al. [8], for example, reported that TPC increased with leaf age, from 65.1 mg GAE/g to 71.5 mg GAE/g. In addition, Annazhifah et al. [4] reported a TPC of smoked, dried coffee leaves of 7.32 ± 0.16 g GAE/100 g dry solid, equivalent to 73.2 ± 1.6 mg GAE/g dry solid (dry basis). Together, these findings underscore that differences in raw material conditions and processing may influence TPC values.

TPC is most commonly determined using the Folin–Ciocalteu colorimetric assay. Results are typically expressed as gallic acid equivalents (GAE) based on a gallic acid calibration curve, and in many applications, the reaction product is measured using a UV–Vis spectrophotometer at 725 nm [9]. Although this method is widely used and can provide reliable results, Folin-based wet-chemistry procedures are relatively time-consuming and labor-intensive, require careful technique and analytical skill, and involve chemical reagents that may generate laboratory waste [10]. Accordingly, faster, non-destructive, and more environmentally friendly approaches are needed to enable efficient TPC screening,

especially when large numbers of samples must be analyzed for quality control or for the development of predictive models.

Near-infrared (NIR) spectroscopy represents one of the most promising approaches to meet these needs. Advances in NIR spectroscopy have created new opportunities for rapid chemical analysis by exploiting spectral patterns, without requiring complex sample preparation or destructive testing. NIR has been increasingly used to assess polyphenol-related attributes in tea and other plant-derived matrices, and several studies have demonstrated practical feasibility for rapid screening [10-15]. Model performance, however, is not uniform; reported outcomes range from good to excellent to moderate depending on matrix characteristics and calibration design [12-15]. This variability is commonly associated with physical-spectral disturbances (e.g., scattering and baseline variation) and sample heterogeneity, which can obscure chemically relevant signals and therefore require appropriate preprocessing [16]. In addition, prediction uncertainty increases when new samples fall outside the calibration domain; thus, explicit applicability domain control and local/domain-adaptive calibration strategies are important for improving transferability across conditions [17, 18]. Careful latent-variable selection and rigorous validation are also essential to prevent underfitting or overfitting and to maintain robust predictive performance [19]. In contrast, NIR studies on coffee leaves have largely emphasized qualitative discrimination, indicating that quantitative TPC modeling in coffee-leaf powder remains underexplored and warrants dedicated development with external validation [20]. Therefore, NIR-based chemometric modeling offers a promising, efficient, and environmentally friendly approach for TPC quality evaluation. This study aimed to evaluate the ability of NIR spectroscopy, combined with PCA for exploration and PLSR for quantitative modeling, to predict TPC in coffee-leaf powder.

2. MATERIALS AND METHODS

2.1 Samples

Robusta coffee leaves (*Coffea canephora* var. *robusta*) processed by traditional smoking were collected from several regions in West Sumatra, Indonesia. The dried leaves were ground using a blender (Philips HR2001; 350 W, 220–240 V, 50–60 Hz) at speed setting 2 for 1 min per cycle, repeated for 3–5 cycles per sample (total grinding time: 3–5 min). The powder was then passed through a 40-mesh sieve. Approximately 15 g of powder was packed into each polypropylene (PP) clip-lock bag (9 × 15 cm), and each bag was treated as an independent sample (n = 90). All bags were fully sealed and stored at ambient room temperature, protected from direct sunlight, for 14 days before NIR scanning and reference chemical analysis. No nitrogen flushing was applied.

2.2 Total phenolic content

TPC was measured using a modified Folin–Ciocalteu assay adapted from earlier studies [10, 21]. Coffee-leaf powder (1.0 g, 40-mesh) was extracted with 10 mL of 70% (v/v) methanol, vortexed, and ultrasonically treated at 70°C for 15 min. After settling, 1.0 mL of supernatant was diluted to 100 mL with distilled water. A 1.0 mL aliquot was then mixed with 1.0 mL

of 10% Folin–Ciocalteu reagent and 1.0 mL of 7.5% Na_2CO_3, vortexed, and incubated for 60 min at room temperature in the dark. Absorbance was read at 725 nm (UV–Vis). TPC (mg GAE/g) was calculated from a gallic acid calibration (10–70 ppm; y = 0.0232x + 0.1227, R^2 = 0.9989) prepared from a 1000 ppm stock; each standard was measured once (n = 1).

2.3 Near-infrared spectroscopy measurements

NIR spectra of ground coffee leaves were recorded using a BUCHI NIRFlex N-500 spectrophotometer (BUCHI Labortechnik AG, Switzerland) controlled with NIRWare software. For each measurement, 15 g of sample was placed in a 10 cm-diameter Petri dish and lightly leveled without pressing, ensuring a flat surface and nearly uniform layer thickness. Spectral acquisition was performed in reflectance mode across 10,000–4,000 cm^{-1} (equivalent to 1,000–2,500 nm) at a resolution of 4 cm^{-1}. Each sample was scanned three times, and the resulting spectra were then used for subsequent chemometric analysis.

2.4 Chemometric analysis

Chemometric processing was carried out using the Unscrambler X software (version 10.3, CAMO ASA, Oslo, Norway). Each sample was scanned three times, and the replicate spectra were averaged to generate one representative spectrum per sample. Before modeling, reflectance data (R) were transformed to absorbance values using A = log(1/R). Spectral exploration and modeling were then performed with principal component analysis (PCA) and partial least squares regression (PLSR).

The 90 samples were first ranked according to their reference TPC values and then split into calibration and external validation sets using an interleaved 2:1 scheme, in which two consecutive samples were assigned to calibration and the next one to validation. This procedure resulted in 60 calibration samples (2/3) and 30 validation samples (1/3). The approach was selected to maintain similar TPC coverage across both subsets while following a split ratio commonly applied in NIR studies [22]. The calibration subset was used to build the model, whereas the validation subset was kept exclusively for external prediction. Within the calibration stage, random-subset cross-validation with 20 segments was applied to determine the optimal number of latent variables [23].

To improve the link between spectral features and reference TPC values, and to minimize multiplicative scattering effects and baseline drift, multiple preprocessing strategies were tested: standard normal variate (SNV), multiplicative scatter correction (MSC), SNV combined with gap-segment derivatives, and MSC combined with gap-segment derivatives. For the gap-segment derivative, the settings were derivative order = 2, gap size = 1, and segment size = 1. Prior to PLSR, all variables were mean-centered [24].

Model performance was assessed using the coefficient of determination for the calibration and prediction sets (R^2_{cal} and R^2_{pred}, respectively), together with the root mean square error for the calibration, cross-validation, and prediction sets ($RMSE_c$, $RMSE_{cv}$, $RMSE_{pred}$, respectively). While R^2 and RMSE are among the most commonly used indicators for evaluating model performance, they should be interpreted with care rather than used as the sole basis for judging model adequacy, since each metric captures only part of the error

structure [25]. For this reason, additional statistical parameters, including the residual predictive deviation (RPD) and range error ratio (RER), were also considered. These metrics were calculated using Eqs. (1)-(4).

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

$$\text{RMSEC, RMSECV and RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$RPD = \frac{SD_{pred}}{RMSEP} \quad (3)$$

$$RER = \frac{\Delta y_{pred}}{RMSEP} \quad (4)$$

where, y_i is the measured (true) value for sample i, $\hat{y}_i$ is the predicted value for sample i, $\bar{y}$ is the mean of the measured values, n is the number of samples, SD is the standard deviation of the prediction set, and Δy is the difference between the maximum and minimum measured values in the prediction set.

Based on R^2, models were categorized into three levels: $0.66 < R^2 < 0.81$ (approximate quantitative prediction), $0.82 < R^2 < 0.90$ (good prediction), and $R^2 > 0.91$ (excellent) [26]. Based on RPD, models were grouped into five levels: < 1.5 (not usable), 1.5–2.0 (possible discrimination between low and high values), 2.0–2.5 (approximate quantitative prediction), 2.5–3.0 (good quantitative prediction), and > 3.0 (excellent prediction) [27]. Furthermore, based on RER, models were classified into four levels: 8–10 (moderately useful), 10–15 (moderately successful), 15–20 (successful), and > 20 (excellent prediction model) [28].

3. RESULTS AND DISCUSSION

3.1 Principal component analysis

An initial analysis over the full wavelength range was performed using PCA, after mean-centering the spectra. According to this study [29], PCA is an effective multivariate tool for reducing the number of variables and summarizing the general characteristics of a dataset through linear combinations of the original variables. In this study, the score plot for the first two principal components (Figure 1) indicated that PC1 and PC2 accounted for 81% and 18% of the variance,

respectively. Collectively, PC1 and PC2 explained 99% of the total data variance, providing an overview of spectral variability without explicitly evaluating the relationship between the spectra and the TPC of coffee-leaf powder.
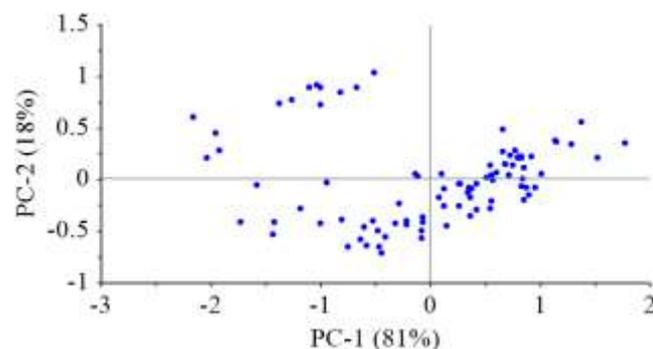

Figure 1. Score plots PC1 vs PC2

3.2 Total phenolic content of the calibration and external validation sets

Table 1 presents the descriptive statistics of TPC in coffee-leaf powder measured using the reference method for the calibration and validation sets. The reference chemical data exhibited a wide range of values, with the calibration samples spanning from 6.31 to 84.40 and the validation samples from 8.15 to 80.70. The mean values of the calibration (37.52) and validation (prediction) sets (37.77) were comparable, indicating that the data split was relatively balanced. The relatively high standard deviations (23.97 for calibration and 23.37 for validation) resulted in coefficients of variation above 60%, reflecting substantial chemical variability among samples.

Substantial chemical variability among samples, together with a wide reference-value range, is advantageous for developing NIR calibrations because it helps construct a calibration set that is more representative of the target population and improves model robustness against biological variability as well as measurement-related effects [30]. In addition, when the range of the validation set falls within the calibration range, external evaluation primarily reflects interpolation; conversely, samples lying outside the calibration domain may lead to more uncertain predictions and should be treated with caution because they fall beyond the model's applicability domain [17]. In such cases, strategies such as local calibration, domain adaptation, or transfer approaches can be considered to accommodate domain shifts and extend model applicability to new conditions [18].

Table 1. Total phenolic content (TPC) of coffee-leaf powder in the calibration and prediction (external validation) sets

Set	n	Range (mg GAE/g sample)	Mean (mg GAE/g sample)	SD (mg GAE/g sample)	CV (%)
Calibration	60	6.31-84.40	37.52	23.97	63.89
Validation	30	8.15-80.70	37.77	23.37	61.87

Note: n: total sample; SD: standard deviation; CV: coefficient of variation (SD/Mean × 100); GAE: gallic acid equivalent.

3.3 Near-infrared spectral characteristics of coffee-leaf powder

Figure 2 shows the raw NIR spectra of coffee-leaf powder recorded over the 1000–2500 nm wavelength range. In general, the raw spectra exhibit absorption patterns typical of plant materials, dominated by overtone and combination bands of functional groups constituting the leaf matrix.

Several absorption features are observed at approximately 1794, 2054, 2205, 2341, 2380, and 2445 nm. According to the study [29], the band near 1794 nm is mainly related to O–H combination vibrations; the band around 2054 nm is associated with N–H combination/amide-related absorptions linked to proteins; and regions above 2200 nm are generally attributed to C–H and N–H combination bands from organic constituents such as cellulose, lipids, and other nitrogen-

containing compounds. However, in the raw spectra, these bands remain relatively broad and overlapping, limiting direct visual resolution of specific chemical features.

Although direct one-to-one assignment of individual NIR bands to specific phenolic molecules in coffee-leaf powder is limited by broad and overlapping overtone or combination absorptions, a chemically plausible interpretation can still be made. Coffee leaves are reported to contain polyphenol-related constituents (including chlorogenic acid derivatives and mangiferin), and variation in these hydroxyl-rich compounds may contribute to O–H-associated absorptions, including the region around ~1794 nm. Meanwhile, the higher-wavelength zone (~2200–2500 nm) contains complex combination bands that may be influenced by phenolic-rich plant matrices together with other constituents [6, 8, 31]. Therefore, in this study, the spectrum–TPC relationship is interpreted as an indirect, matrix-level chemometric association rather than a unique fingerprint of a single compound, consistent with previous NIR-based phenolic prediction studies in tea or plant matrices [10, 11, 14, 32]. Their contribution to TPC prediction is further examined in Section 3.4 using the regression-coefficient profile of the best PLSR model.

The application of several preprocessing approaches prior to chemometric modeling of the raw NIR spectra is shown in Figure 3. Both SNV and MSC markedly reduced baseline variation and scattering effects compared with the raw data. SNV produced strongly normalized spectra, improving inter-sample alignment but sacrificing the direct absorbance scale. In contrast, MSC preserved the relative absorbance magnitude while effectively correcting global scattering effects. This difference can influence subsequent chemometric performance, because SNV emphasizes relative spectral variation, whereas MSC retains intensity information that may be chemically meaningful.
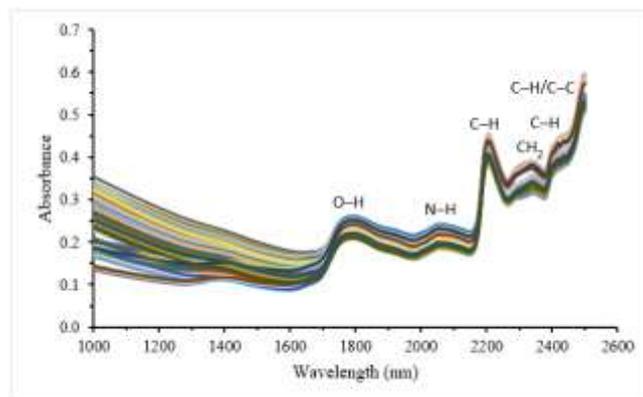

Figure 2. Raw near-infrared (NIR) absorbance spectra of coffee-leaf powder samples in the range 1000–2500 nm
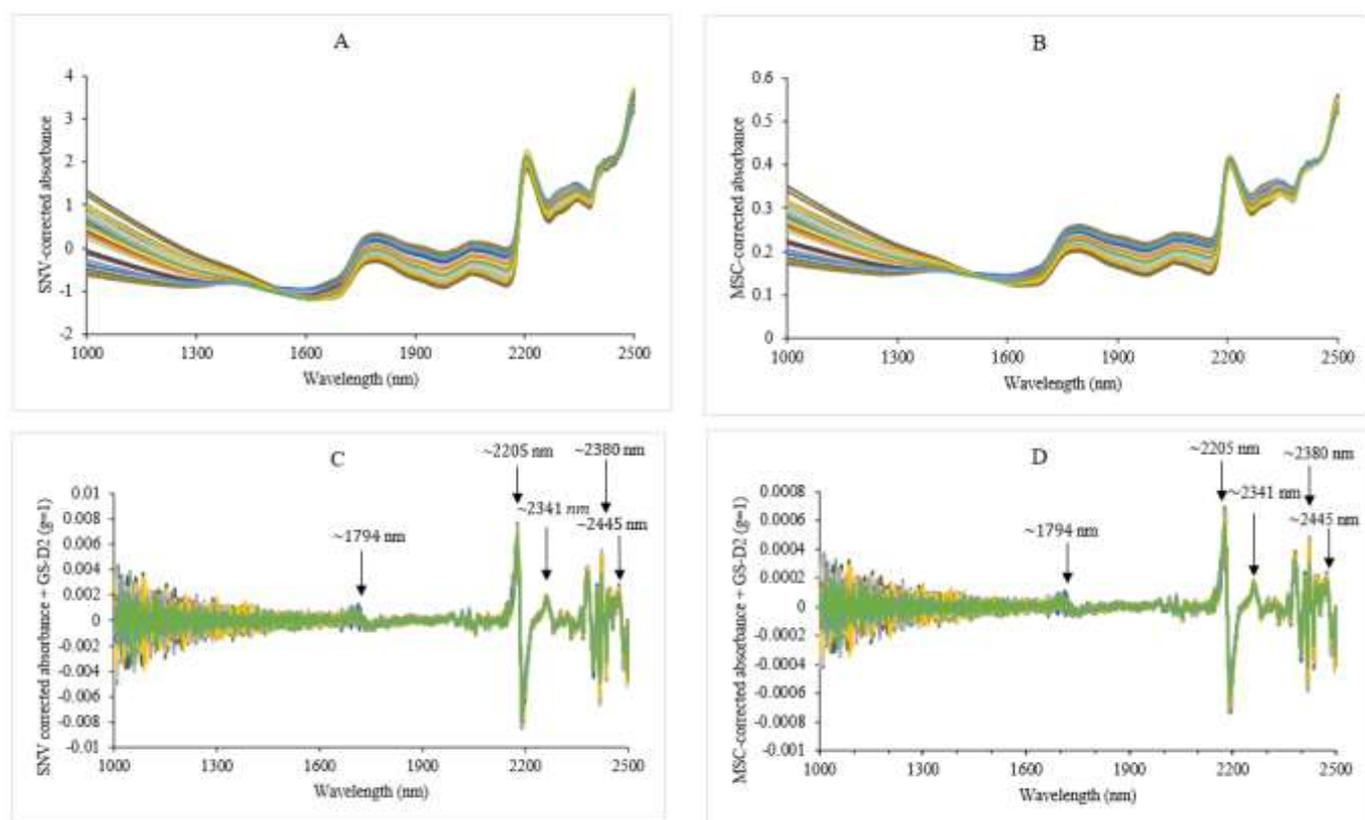

Figure 3. Preprocessed near-infrared (NIR) spectra of coffee-leaf powder over 1000–2500 nm: (A) SNV, (B) MSC, (C) SNV followed by gap-segment second derivative (GS-D2), and (D) MSC followed by gap-segment second derivative (GS-D2)
In the derivative-treated spectra (C and D), previously overlapping absorptions are more clearly resolved, with local features around ~1794, ~2205, ~2341, ~2380, and ~2445 nm.

To further enhance subtle spectral features and remove residual baseline effects, a second-derivative transformation using a gap-segment algorithm was applied to the SNV- and MSC-corrected spectra. This approach sharpened overlapping absorption bands and improved spectral resolution, particularly in regions that are likely to carry chemically relevant information. After second-order gap-segment derivative preprocessing, overlapping bands were

better separated, particularly in the higher-wavelength region. Clear local features were observed at approximately ~1794, ~2205, ~2341, ~2380, and ~2445 nm, supporting a more specific spectral interpretation than that obtained from raw or pre-corrected spectra. A comparative evaluation of the four preprocessing methods enabled an objective selection of the optimal approach based on predictive performance rather than visual appearance alone.

3.4 Partial least squares regression calibration model for total phenolic content prediction

Table 2 summarizes the performance of PLSR models for predicting the TPC of coffee-leaf powder from NIR spectra by comparing several spectral preprocessing schemes (Original, SNV, MSC, and SNV/MSC combined with a second-order gap-segment derivative). For the Original/SNV/MSC spectra, the models required 11 latent variables (LVs), whereas only 4 LVs were sufficient when the second-order derivative (with the gap-segment algorithm) was applied. This marked reduction in LV suggests that the derivative preprocessing successfully extracted cleaner and more relevant information, allowing the relationship between the spectra and TPC to be modeled more parsimoniously. Selecting an appropriate number of LVs is a critical step in model development because it directly affects performance: too few LVs may lead to underfitting, whereas too many may cause overfitting and ultimately degrade the model's future predictive capability [19].

Table 2. Calibration model performance

Spectra Pre-Processing	LVs	Calibration				Validation				
		R^2	RMSEC	RMSECV	SEC	R^2	RMSEP	SEP	RPD	RER
Original	11	0.96	4.65	8.15	4.69	0.93	6.39	6.5	3.66	11.35
SNV	11	0.98	3.73	6.93	3.76	0.93	6.22	6.4	3.76	11.66
MSC	11	0.98	3.72	6.55	3.76	0.91	7.12	7.31	3.28	10.19
SNV_2nd Gap Segment Derivatives	4	0.97	4.15	8.04	4.18	0.95	5.14	5.77	4.55	14.11
MSC_2nd Gap Segment Derivatives	**4**	**0.97**	**4.14**	**8.03**	**4.18**	**0.97**	**4.96**	**5.53**	**4.71**	**14.63**

Note: LVs: latent variables; RMSEC: root mean squared error of calibration; RMSECV: root mean squared error of cross-validation; RMSEP: root mean squared error of prediction; RPD: residual predictive deviation; RER: range error ratio; SNV: standard normal variate, MSC: multiplicative scatter correction.

The raw-spectra (Original) model achieved an external validation R^2pred of 0.93, with an RMSEP of 6.39 and an RPD of 3.66. This performance falls within the excellent category for quantitative prediction (RPD > 3) [27], indicating that TPC-related chemical information is indeed captured by the spectra, although residual physical effects (e.g., scattering and baseline variation) may still limit precision.

Applying SNV led to a moderate improvement in performance (RMSEP decreased from 6.39 to 6.22; RPD increased from 3.66 to 3.76). This improvement is consistent with the purpose of SNV, which is to suppress multiplicative variation caused by scattering and optical pathlength/thickness effects [33]. In contrast, using MSC alone resulted in a higher RMSEP (7.12) and a lower RPD (3.28). This pattern suggests that a scattering-correction approach that is not well aligned with the dominant sources of variation and the analytical objective may confound the target information with unwanted variability, and may even attenuate relevant signals in certain spectral regions; the magnitude of this effect depends on sample characteristics and measurement conditions [16].

The most pronounced improvement was observed when scatter correction was combined with the second derivative (gap/segment derivatives). The SNV + second-derivative model reduced RMSEP to 5.14 and increased RPD to 4.55, while requiring only 4 LVs. The best model in this study was obtained using MSC combined with a second-order derivative, achieving an external validation R^2 of 0.97, RMSEP of 4.96, SEP of 5.53, RPD of 4.71, and RER of 14.63. Relative to the raw-spectra model, RMSEP decreased by ~22%, and the RPD/RER indices increased by ~29%, while model complexity was simultaneously reduced (LV 11 to LV 4). From an interpretive standpoint, these results are consistent with the nature of solid samples, which are prone to scattering-related variability and therefore benefit from scatter correction. In addition, second-derivative preprocessing is commonly applied in NIR spectroscopy to reduce offset and slope effects and to minimize baseline variation, thereby enhancing chemically driven variation in the model [16].

The best performance achieved in this study (RPD of 4.71; RER of 14.63; external validation R^2 of 0.97) indicates a very strong predictive capability compared with several NIR/Vis–NIR–based TPC studies on other commodities. For example, in rocket leaves (*Eruca sativa*), a Vis–NIR model for TPC was reported to reach a validation R^2 of 0.84, RPD of 3.27, and RER of 12.5, which is commonly considered good for quantitative prediction [14]. For berry extracts, reported TPC prediction performance also varies, typically with $R^2 > 0.84$ and RPD values around 1.8–3.1, depending on the berry type and the spectral configuration used [12].

In other applications, performance can be lower. Rouxinol et al. [13], for instance, reported TPC prediction for red wine using a portable NIR device with $R^2 = 0.71$ and RPD = 1.6, which they interpreted as very low quantitative capability to borderline performance. Collectively, these comparisons reinforce that successful prediction of TPC/phenolics is strongly influenced by the sample matrix, the breadth of variability in the reference range, and the design of the calibration and validation sets. In line with this, more recent studies on other food materials also suggest that TPC prediction often represents a moderately challenging task; Zheng et al. [15] reported prediction R^2 values of approximately 0.79 for some of the tested objects.

Beyond cross-matrix comparisons, we added direct tea-matrix benchmarks for NIR-based TPC or polyphenol prediction. Liu et al. [10] reported very high performance in Fuzhuan tea (R^2 in prediction set = 0.9996; RMSEP ≈ 0.0611), while Yin et al. [34] reported robust fresh-tea performance (best RPD = 2.9226; overall RPD = 2.07–4.06). Chanda et al. [11] further supported inward tea-leaf polyphenol prediction using Grey Wolf Optimization, and Turgut et al. [35] showed broad PLSR performance ranges in black tea (sensory: R^2cv 0.83–0.97, RPDcv 2.47–5.79; analytical: R^2cv 0.66–0.89, RPDcv 1.72–3.08). Thus, performance differences across studies are better interpreted as method–matrix interactions,

not algorithm effects alone.

Performance in NIR-based TPC prediction is strongly influenced by matrix characteristics and modeling strategy [10-15]. Across studies, relative differences may be associated with matrix composition and sample form or homogenization, preprocessing strategy (including derivative/scatter correction), and calibration/validation design [12-16]. Therefore, cross-study performance should be interpreted mainly as a method–matrix interaction rather than an algorithm effect alone [10-15]. Finally, the best model from the present study is illustrated in Figure 4, which plots reference versus predicted values for the TPC of coffee-leaf powder based on the optimal calibration model.
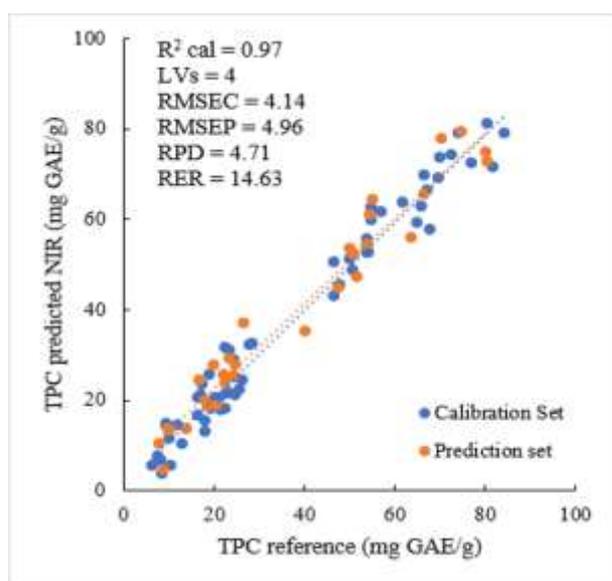

Figure 4. Scatter plot of reference versus predicted total phenolic content (TPC) for coffee-leaf powder using the best PLSR calibration model
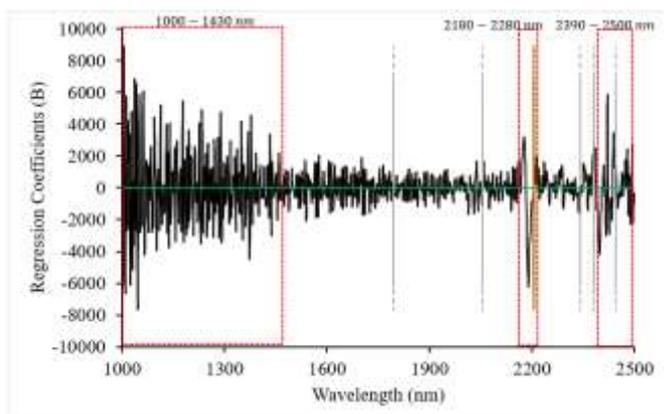

Figure 5. Regression coefficients of the best partial least squares regression (PLSR) model for total phenolic content (TPC) prediction (1000–2500 nm)
High-contribution regions (high |B|) are highlighted at approximately 1000–1430 nm, 2150–2280 nm, and 2341–2500 nm.

Figure 5 shows the regression-coefficient profile of the best PLSR model for TPC prediction across 1000–2500 nm. The model depends on multiple spectral intervals rather than a single isolated wavelength. High absolute coefficients (|B|) are mainly observed at 1000–1430 nm, 2180–2280 nm, and 2390–2500 nm, indicating that these regions contribute most

strongly to prediction. The two higher-wavelength intervals are consistent with the spectral characterization in Section 3.3 (> 2200 nm), suggesting that the model captures chemically relevant information from the coffee-leaf matrix. The coefficient sign (positive or negative) indicates the direction of influence on the predicted TPC, while variable importance is interpreted from the coefficient magnitude.

4. CONCLUSIONS

NIR spectroscopy combined with PLSR enabled highly accurate prediction of TPC in coffee-leaf powder under external validation, indicating that polyphenol-related chemical information is effectively captured in the material's spectral patterns. Among the evaluated preprocessing schemes, MSC coupled with a second-order gap-segment derivative delivered the strongest performance while substantially reducing the required number of latent variables, resulting in a more parsimonious model without compromising accuracy. The wide TPC variability observed in both the calibration and validation sets further supported model robustness by reflecting the inherent heterogeneity of the material.

From a practical perspective, these findings support the use of NIR as a rapid, non-destructive, and more environmentally friendly method for quality screening of coffee-leaf powder (e.g., for raw-material selection, process standardization, or routine quality control). Future improvements may be achieved by expanding sample coverage (origin, season, and processing variations), standardizing the reporting basis (e.g., dry basis when moisture content is available), and testing model transferability across time and/or instruments to ensure more stable field deployment.

REFERENCES

[1] Sutrisno, A., Wahyuni, E., Agang, M.W., Hartono, T.T., Sayaza, M.D., Santoso, D., Titing, D., Kusnadi, E., Novita, E., Pramulya, R., Rahmah, D.M. (2025). Sustainability evaluation of robusta coffee farming in Malinau Regency using the sustainable livelihood framework. Organic Farming, 11(2): 72-89. https://doi.org/10.56578/of110201

[2] Chen, X., Ma, Z., Kitts, D.D. (2018). Effects of processing method and age of leaves on phytochemical profiles and bioactivity of coffee leaves. Food Chemistry, 249: 143-153. https://doi.org/10.1016/j.foodchem.2017.12.073

[3] Novita, R., Kasim, A., Anggraini, T., Putra, D.P. (2018). Kahwa daun: Traditional knowledge of a coffee leaf herbal tea from West Sumatera, Indonesia. Journal of Ethnic Foods, 5(4): 286-291. https://doi.org/10.1016/j.jef.2018.11.005

[4] Annazhifah, N., Syamsir, E., Herawati, D. (2024).

Stability of phenolic antioxidants in coffee leaf beverage during pasteurization. Food Research, 8(5): 282-288. https://doi.org/10.26656/fr.2017.8(5).512

[5] Fibrianto, K., Bimo, I.A., Wulandari, E.S., Hendrawan, Y. (2025). Functional and relaxing properties of coffee leaf tea: An integrative food cognition approach. Applied Food Research, 5(2): 101141. https://doi.org/10.1016/j.afres.2025.101141

[6] Arifan, F., Broto, W., Supriyo, E., Faisal, M.M., Wardani, O.K., Sapatra, E.F. (2023). Characterization of physical and chemical properties of functional beverages of robusta coffee leaf herbal tea with red ginger-enriched green tea technique. Materials Today: Proceedings, 87: 350-354. https://doi.org/10.1016/j.matpr.2023.03.622

[7] Monteiro, Â., Colomban, S., Azinheira, H.G., Guerra-Guimarães, L., Do Céu Silva, M., Navarini, L., Resmini, M. (2019). Dietary antioxidants in coffee leaves: Impact of botanical origin and maturity on chlorogenic acids and xanthones. Antioxidants, 9(6): 1-16. https://doi.org/10.3390/antiox9010006

[8] Acidri, R., Sawai, Y., Sugimoto, Y., Handa, T., et al. (2020). Phytochemical profile and antioxidant capacity of coffee plant organs compared to green and roasted coffee beans. Antioxidants, 9(2): 93. https://doi.org/10.3390/antiox9020093

[9] Jeong, S., Kim, S.Y., Myeong, H., Lim, E.K., et al. (2024). Microbead-based colorimetric and portable sensors for polyphenol detection. ACS OMEGA, 9: 36531-36539. https://doi.org/10.1021/acsomega.4c04523

[10] Liu, J.X., Xin, J.Y., Gao, T.T., Li, F.L., Tian, X. (2022). Effect of variable selection and rapid determination of total tea polyphenols contents in Fuzhuan tea by near-infrared spectroscopy. CyTA - Journal of Food, 20(1): 236-243. https://doi.org/10.1080/19476337.2022.2128429

[11] Chanda, S., Sing, D., Majumder, S., Nag, S., et al. (2017). NIR spectroscopy with grey wolf optimization algorithm for prediction of polyphenol content in inward tea leaves. In 2017 IEEE Calcutta Conference (CALCON), Kolkata, India, pp. 392-396. https://doi.org/10.1109/CALCON.2017.8280762

[12] Kljusurić, J.G., Mihalev, K., Bečić, I., Polović, I., Georgieva, M., Djaković, S., Kurtanjek, Ž. (2016). Near-infrared spectroscopic analysis of total phenolic content and antioxidant activity of berry fruits. Food Technology and Biotechnology, 54(2): 236-242. https://doi.org/10.17113/ftb.54.02.16.4095

[13] Rouxinol, M.I., Martins, M.R., Murta, G.C., Barroso, J.M., Rato, A.E. (2022). Quality assessment of red wine grapes through NIR spectroscopy. Agronomy, 12(3): 637. https://doi.org/10.3390/agronomy12030637

[14] Toledo-Martín, E.M., Font, R., Obregón-Cano, S., De Haro-Bailón, A., Villatoro-Pulido, M., Del Río-Celestino, M. (2017). Rapid and cost-effective quantification of glucosinolates and total phenolic content in rocket leaves by visible/near-infrared spectroscopy. Molecules, 22(5): 851. https://doi.org/10.3390/molecules22050851

[15] Zheng, C., Li, J., Liu, H., Wang, Y. (2024). Rapid and non-invasive estimation of total phenol content and species identification in dried wild edible bolete using FT-NIR spectroscopy. Arabian Journal of Chemistry, 17(12): 106011. https://doi.org/10.1016/j.arabjc.2024.106011

[16] Rinnan, A. (2014). Pre-processing in vibrational spectroscopy, a when, why and how. Analytical Methods, 6(18): 7124-7129. https://doi.org/10.1039/c3ay42270d

[17] Rodríguez-Barrios, M.S., Ferre, J., Larrechi, M.S., Ruiz, E. (2024). Chemometrics and intelligent laboratory systems applicability domain of a calibration model based on neural networks and infrared spectroscopy. Chemometrics and Intelligent Laboratory Systems, 254: 105242. https://doi.org/10.1016/j.chemolab.2024.105242

[18] Yang, X.P., Yang, F.Y., Lesnoff, M., Berzaghi, P., Ferragina, A. (2024). Diverse local calibration approaches for chemometric predictive analysis of large near-infrared spectroscopy (NIRS) multi-product datasets. Chemometrics and Intelligent Laboratory Systems, 251: 105173. https://doi.org/10.1016/j.chemolab.2024.105173

[19] Gowen, A.A., Downey, G., Esquerre, C., Donnell, C.P.O. (2011). Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. Journal of Chemometrics, 25(7): 375-381. https://doi.org/10.1002/cem.1349

[20] Mees, C., Souard, F., Delporte, C., Deconinck, E., et al. (2018). Identification of coffee leaves using FT-NIR spectroscopy and SIMCA. Talanta, 177: 4-11. https://doi.org/10.1016/j.talanta.2017.09.056

[21] Ivanović, S., Avramović, N., Dojčinović, B., Trifunović, S., Novaković, M., Tešević, V., Mandić, B. (2020). Chemical composition, total phenols and flavonoids contents and antioxidant activity as nutritive potential of roasted hazelnut skins (Corylus avellana L.). Foods, 9(4): 430. https://doi.org/10.3390/foods9040430

[22] Díaz-maroto, I.J., Soledad, M.P., Díaz-maroto, M.C., Alarc. (2023). Rapid and non-invasive estimation of total polyphenol content and antioxidant activity of natural corks by NIR spectroscopy and multivariate analysis. Food Packaging and Shelf Life, 38: 101099. https://doi.org/10.1016/j.fpsl.2023.101099

[23] Páscoa, R.N.M.J., Magalhães, L.M., Lopes, J.A. (2013). FT-NIR spectroscopy as a tool for valorization of spent coffee grounds: Application to assessment of antioxidant properties. Food Research International, 51(2): 579-586. https://doi.org/10.1016/j.foodres.2013.01.035

[24] Zhu, M.T, Long, Y., Chen, Y., Huang, Y.S., et al. (2021). Fast determination of lipid and protein content in green coffee beans from different origins using NIR spectroscopy and chemometrics. Journal of Food Composition and Analysis, 102: 104055. https://doi.org/10.1016/j.jfca.2021.104055

[25] Santhosh, C.S., Umesh, K.K., Hemanth, V., Narendra, K. (2025). Forecasting yield of coffee crop varieties C × R, Sln3 and Sln5B: A stochastic machine learning model based on agro-ecological factors using multivariate feature selection approach. Organic Farming, 11(3): 203-226. https://doi.org/10.56578/of110305

[26] Saha, U., Endale, D., Tillman, P.G., Johnson, W.C., et al. (2017). Analysis of various quality attributes of sunflower and soybean plants by near infrared reflectance spectroscopy: Development and validation calibration models. American Journal of Analytical Chemistry, 8(7): 462-492. https://doi.org/10.4236/ajac.2017.87035

[27] Mouazen, A.M., Saeys, W., Xing, J., De Baerdemaeker, J., Ramon, H. (2005). Near infrared spectroscopy for agricultural materials: An instrument comparison. Journal of Near Infrared Spectroscopy, 13(2): 87-97. https://doi.org/10.1255/jnirs.461

[28] Widyaningrum, W., Purwanto, Y.A., Widodo, S., Supijatno, S., Iriani, E.S. (2025). Portable near-infrared spectroscopy and support vector regression for fast quality evaluation of Vanilla (Vanilla planifolia). Jurnal Teknik Pertanian Lampung, 14(2): 515-526. https://doi.org/10.23960/jtep-l.v14i2.515-526

[29] Araújo, C. da S., Macedo, L.L., Vimercati, W.C., Saraiva, S.H. (2021). Spectroscopy technique applied to estimate sensory parameters and quantification of total phenolic compounds in coffee. Food Analytical Methods, 14(9): 1943-1952. https://doi.org/10.1007/s12161-021-02025-0

[30] Zhang, X., Yang, J. (2024). Advanced chemometrics toward robust spectral analysis for fruit quality evaluation. Trends in Food Science & Technology, 150: 104612. https://doi.org/10.1016/j.tifs.2024.104612

[31] Burns, D.A., Ciurczak, E.W. (2008). Handbook of Near-Infra-Red Analysis (3rd ed.). CRC Press, Boca Raton, Florida.

[32] Revilla, I., Jiménez, M.H., Martínez-Martín, I., Valderrama, P., Rodríguez-Fernández, M., Vivar-Quintana, A.M. (2024). The potential use of near infrared spectroscopy (NIRS) to determine the heavy metals and the percentage of blends in tea. Foods, 13(3): 450. https://doi.org/10.3390/foods13030450

[33] Fan, C., Liu, Y., Cui, T., Qiao, M., Yu, Y., Xie, W., Huang, Y. (2024). Quantitative prediction of protein content in corn kernel based on near-infrared spectroscopy. Foods, 13(24): 4173. https://doi.org/10.3390/foods13244173

[34] Yin, X., Xiao, Y.B., Li, J., Pei, Y.Q., Shen, Y.Y., Wang, X.Y. (2025). Optimized NIRS-machine learning framework for rapid multi-trait quality assessment of fresh tea leaves. LWT, 237: 118727. https://doi.org/10.1016/j.lwt.2025.118727

[35] Turgut, S.S., Entrenas, J.A., Taşkın, E., Garrido-Varo, A., Pérez-Marín, D. (2022). Estimation of the sensory properties of black tea samples using non-destructive near-infrared spectroscopy sensors. Food Control, 142: 109260. https://doi.org/10.1016/j.foodcont.2022.109260