AIMS Agriculture and Food, 11(3): 510–530.
DOI: 10.3934/agrfood.2026026
Received: 07 January 2026
Revised: 27 May 2026
Accepted: 09 July 2026
Published: 15 July 2026

https://www.aimspress.com/journal/agriculture

Research article

Spectral difference-based adjustment of stem near-infrared spectra for indirect prediction of durian pulp dry matter and maturity

Kaewkarn Phuangsombut, Arthit Phuangsombut, Sirinad Noypitak and Anupun Terdwongworakul*

Department of Agricultural Engineering, Faculty of Engineering at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, 73140, Thailand

* **Correspondence:** Email: fengant@ku.ac.th; Tel: +66867751723.

Abstract: Indirect prediction of pulp dry matter content (DM) was investigated using near-infrared spectra acquired from fruit stems. The spectral difference between stem and pulp spectra, together with pulp DM, was used to adjust the stem spectra, thereby integrating pulp-related information into the stem spectral data. Partial least squares regression models were developed using unprocessed, preprocessed, and adjusted stem spectra for comparison. The adjusted spectra model achieved a prediction correlation coefficient of 0.828 and a root mean square error of prediction of 1.630%, marginally outperforming the unprocessed and preprocessed models. The best preprocessed model and the adjusted spectra model were subsequently used to build support vector machine classifiers. In binary classification, the adjusted spectra model showed slightly improved performance on the test set, with average accuracy, precision, and recall values of 0.882, 0.833, and 0.923, respectively. This study should be viewed as a proof of concept. Future validation using independent samples from different seasons, orchards, production lots, and environmental conditions is necessary to confirm the robustness and generalizability of the approach.

Keywords: durian; maturity; adjustment; near-infrared spectroscopy; stem

1. Introduction

Durian (*Durio zibethinus* Murray) is one of Thailand's most economically important fruits.

Among the Association of Southeast Asian Nations countries, Thailand is recognized as the leading producer of durian and a major exporter to the international market [1]. Durian production in Thailand is estimated to reach 1.60 million tons in 2026 [2]. Nutritionally, the fruit is highly energy-dense, providing around 147 kcal per 100 g of pulp and is a rich source of carbohydrates, dietary fiber, potassium, vitamin C, thiamine (B1), and vitamin B6 [3]. As a seasonal crop, peak supply in Thailand occurs between June and August. Generally, the popular Monthong cultivar takes 120–135 days from anthesis to attain horticultural maturity suitable for harvesting. Harvest timing is essential for durian quality, as fruits harvested too early fail to ripen fully, resulting in poor flavor and an unsatisfactory eating experience [4].

Demand for durian is consistently high at the beginning of the harvest season, ensuring premium prices for exporters. However, in pursuit of these high early-season market prices, growers may be tempted to harvest fruit as soon as or before it reaches physiological maturity. This poses a major challenge for exporters, as immature and mature fruits are often harvested simultaneously. To maintain high-quality supply and protect market reputation, it is essential to select fruit at the proper stage of maturation, and sorting immature fruit from mature fruit is challenging, even for experienced individuals. Therefore, it would be highly beneficial to develop a rapid and non-destructive technique to accurately determine the degree of maturation.

Several criteria are used to assess durian fruit maturity, including the number of days from full bloom, the color of the spine tip, the elasticity and arrangement of the spines, the intensity of the emitted odor, changes in the fruit stalk, the percentage of dry matter, and the sound produced when tapped [4,5]. Only experienced and highly skilled farmers can effectively utilize these maturity indicators. To reduce reliance on manual sorting, the dry matter (DM) content in durian pulp has been identified as a key indicator of maturity because it accumulates consistently during the stationary phase of fruit growth, which correlates with durian maturity. In contrast, other chemical components, such as fat, total sugar, total non-structural carbohydrates, total soluble solids, and β-carotene, fluctuate throughout development [5]. In Thailand, the DM content is recognized as the standard for assessing durian maturity, as officially defined in the Thai Agricultural Commodity and Food Standard (TACFS 3–2003), according to the National Bureau of Agricultural Commodity and Food Standards [6]. However, measuring the DM content requires a destructive testing process.

Near-infrared spectroscopy (NIRS) has been widely used for classifying and quantifying durians. NIRS can be applied to analyze pulp spectra for the development of predictive models for the DM content. For example, Onsawai et al. [7] constructed a prediction model using the 800–2500 nm range, achieving a coefficient of determination (Rp^2) of 0.89 and a root mean square error of prediction (RMSEP) of 3.60%. Pokhrel et al. [8] applied NIRS to durian pulp in the 300–1160 nm range, achieving 85.3% accuracy in classifying durian pulp into three categories based on the DM content. Additionally, Puttipipatkajorn et al. [9] used a miniature NIR spectrometer with an artificial neural network (ANN) algorithm to predict the DM content in durian pulp, resulting in Rp^2 and RMSEP values of 0.84 and 1.71%, respectively. Saenphon et al. [10] applied long-wavelength NIR reflectance to scan durian pulp on a moving conveyor belt. Using linear discriminant analysis, they achieved a two-group classification accuracy of 94.2%. Additionally, their DM prediction model, which incorporated selected wavelengths using a genetic algorithm, achieved Rp^2 and SEP values of 0.85 and 4.5%, respectively. Similarly, in-line NIRS measurement of durian pulp on a conveyor has been investigated for in-line prediction of pulp DM [11]. The NIR spectral data in the range 860–1760 nm were used to develop the prediction model with deep neural network regression. High accuracy of prediction was reported, with Rp^2 and RMSEP values of 0.94 and 3.13%, respectively.

A major challenge in the quantitative and qualitative analyses of durian is using NIRS on intact fruits based on a practical, non-destructive measurement system. Indirect prediction has been investigated based on the pulp DM content using NIR information from other parts of the fruit. For example, Ditcharoen et al. [12] created the classification model by fusing NIR spectra from the rind and stem, reporting a classification accuracy of 100% for four distinct maturity stages. Independently, Puttipipatkajorn et al. [9] explored spectral fusion for DM prediction using a miniature NIR spectrometer. By developing an ANN model based on the rind and stem spectra in the 900–1700 nm range, they achieved satisfactory predictive performance, with Rp^2 and RMSEP values of 0.67 and 2.36%, respectively. Imai et al. [13] used portable NIR spectroscopy based on the peduncle to evaluate the pulp DM content. The models based on NIR information of the dried peduncle powder achieved Rp2 > 0.78 for the prediction of slow-maturing fruits. These researchers focused primarily on directly utilizing spectra acquired from specific tissues or combining spectra from multiple tissues. To the best of our knowledge, no other researcher has attempted to improve the predictive capability of stem spectra through a spectral adjustment procedure derived from the relationship between stem spectra, pulp spectra, and pulp dry matter content.

In this study, we investigated the application of NIRS for the indirect prediction of the DM content in durian pulp and for the classification of fruit maturity using stem spectra. In addition, a method was developed to enhance stem spectra by incorporating spectral differences between the pulp and stem together with the DM value, thereby embedding pulp-related information into the stem spectra. Then, the predictive and classification performance of models based on the original and preprocessed stem spectra was compared to those using the adjusted stem spectra.

2. Materials and methods

2.1. Preparation of durian samples

Fresh ‘Monthong’ durians of export-grade size were harvested from an orchard in Chanthaburi province, eastern Thailand. Flowers on selected trees were tagged at full bloom to track fruit development. Several trees were selected to obtain a sufficient number of durian samples, and the fruits were harvested between 97 and 132 days after anthesis, representing stages ranging from immature to overmature. On each harvest date, fruits were collected, transported to the laboratory on the same day, and stored overnight for acclimatization at 25°C. Measurements were conducted the following day.

2.2. Spectral acquisition

The characteristics of the durian stem are associated with fruit maturity [5]. Therefore, in this study, the NIR spectra of the stem were acquired using a spectrophotometer (Model IA500; Bran+Luebbe GmbH; Norderstedt, Germany) to develop a model for the indirect evaluation of pulp quality. The system operated in diffuse reflectance mode using a wavelength range of 1100–2500 nm, with data collected at 2 nm intervals, capturing signals from the stem of the prepared durian samples.

The durian stem can be divided naturally into two parts based on the abscission zone (Figure 1):

The upper part connects to the tree, while the lower part is attached to the fruit. The stem measurements were taken from the surfaces on two opposite sides of the lower part of the stem. These surfaces were chosen to align with two fertile locules.

The scanning area was prepared by carefully peeling a section of the lower stem (approximately 10 mm × 20 mm) near the abscission zone (approximately 0.5 mm thick). Because the sample holder could not accommodate the entire fruit, the prepared lower stem was detached from the fruit and the upper stem. Then, the stem was positioned horizontally in the sample holder (Figure 2). The holder was adjusted using a knob to align the prepared stem area with the focal position for NIR diffuse reflectance scanning. In the focal position, the distance between the light source and the sample was fixed, compensating for variations in sample thickness. Each prepared surface of the stem was scanned at two adjacent locations. The average of the two spectra was used as the stem spectral data for further analyses.

In addition, NIR spectra of the durian pulp were acquired for the spectral adjustment procedure. Measurements were taken from the middle portions of the pulp in two fertile locules of each fruit. The resulting spectra were averaged to represent the pulp spectrum of that fruit.

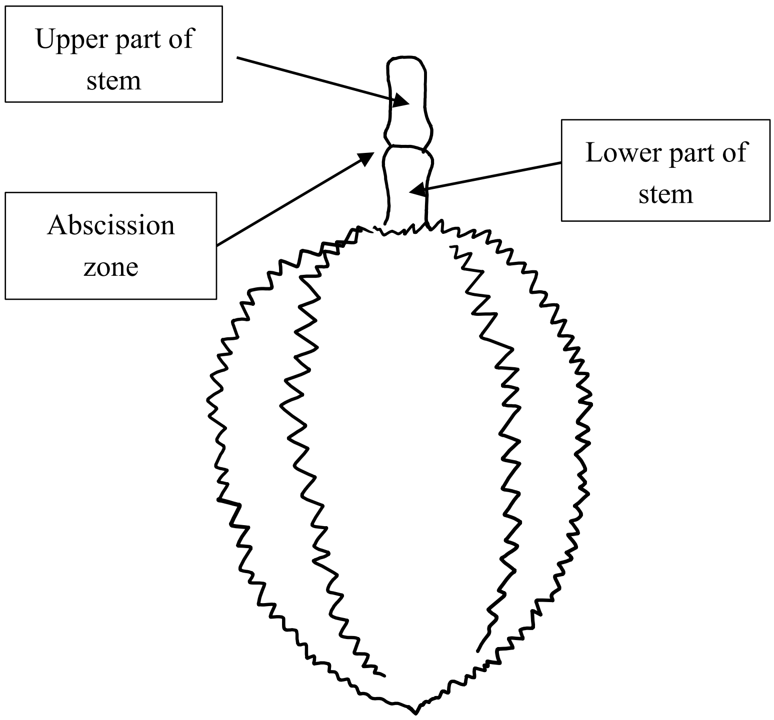

Figure 1. Durian fruit showing the upper and lower parts of the stem, separated by the abscission zone.

Figure 2. Placement of the durian stem, separated from the fruit, in the sample drawer of the spectrometer.

2.3. Determination of pulp dry matter content

During the growth and maturation of 'Monthong' durian, the DM content of the pulp is the most reliable indicator of maturity [14]. Therefore, we used the DM content as the reference standard for assessing maturity.

To determine the DM content, pulp was collected from the middle segments of each locule, consistent with the NIR study, on the pulp and then cut into small pieces and thoroughly mixed. A 20 g sample of the mixed pulp was dried in a hot-air oven at 70°C for 48 hours, following the method described by Sangwanangkul and Siriphanich [14]. The DM percentage was calculated based on the formula (w2 / w1) × 100, where w1 is the fresh weight of the pulp prior to drying and w2 is the weight following complete drying.

2.4. Development of DM prediction model based on stem NIR spectra

2.4.1. Preparation of NIR spectra

The original NIR spectra of the stem and the preprocessed NIR spectra were prepared to reduce the scattering effect in the spectra. The preprocessing techniques consisted of the standard normal variate (SNV), first derivative (D1), and second derivative (D2). In addition, the spectra were preprocessed twice using SNV and D1 (SNV-D1), and SNV and D2 (SNV-D2).

2.4.2. Adjustment of stem spectra using spectral difference and pulp dry matter

The adjustment was empirically achieved by integrating the spectral difference information of the pulp and pulp dry matter into the stem spectra, as outlined in the following steps (Figure 3):
1. Preprocessing: The stem spectra and the pulp spectra were preprocessed using the SNV to reduce the scattering effects.

2. Spectral difference calculation: The spectral difference (SDPS) was obtained by subtracting the SNV stem spectra (SNVSS) from the SNV pulp spectra (SNVPS), as shown in equation (1):

$$SDPS = SNVPS - SNVSS \quad (1)$$

3. Relationship between spectral difference and DM: Chemical changes in the stem are influenced by fruit maturity, which is associated with an increase in the pulp DM [14]. Therefore, the spectral difference, which carried additional pulp-related information into the stem spectra, was hypothesized to correlate with the DM content. At each wavelength, a separate regression model was developed to describe the relationship between the spectral difference and DM content using samples from the calibration set, as expressed in equation (2):

$$DM(\lambda_i) = b_{0i} + b_{1i} \times SDPS(\lambda_i) \quad (2)$$

where $SDPS(\lambda_i)$ is the SDPS at wavelength i, b_{0i} is the intercept at wavelength i, and b_{1i} is the regression coefficient at wavelength i.

4. PLSR model development for DM prediction: Next, a partial least squares regression (PLSR) model was developed to predict DM content using the SNV-preprocessed stem spectra of the calibration set samples as predictor variables.

5. Computation of enhanced DM normalized by SDPS: The term b_{1i} in equation (2) represents the DM value normalized by SDPS for each wavelength (as shown in equation (3)). The PLSR model for DM prediction and the simple regression sets relating DM and SDPS were applied sequentially to obtain an enhanced DM value normalized by SDPS. For each sample, the DM was first predicted using the PLSR model with the SNV stem spectra as input. Then, at each wavelength, the predicted DM of the sample was substituted into equation (2). Based on equations (2) and (3), the calculated DM was an enhanced DM value normalized by the spectral difference. This procedure was repeated across all wavelengths, with the complete normalized DM being obtained by combining the normalized DM values of all wavelengths.

$$b_{1i} = DM(\lambda_i) - \frac{b_{0i}}{SDPS(\lambda_i)} \quad (3)$$

6. Adjustment of stem spectra: Finally, the SNV stem spectra were adjusted (ASS) using the predicted DM (PDM) according to equation (4):

$$ASS = SNVSS + PDM \quad (4)$$

The resulting ASS represented the adjusted stem spectra that included the enhanced DM values normalized by the spectral difference.

The simple regression models for wavelength-specific enhanced DM and the PLSR model for DM prediction could be applied to adjust the stem spectra of unknown samples in the future application by following steps 5 and 6. The only required input from a new sample was its SNV stem spectra.

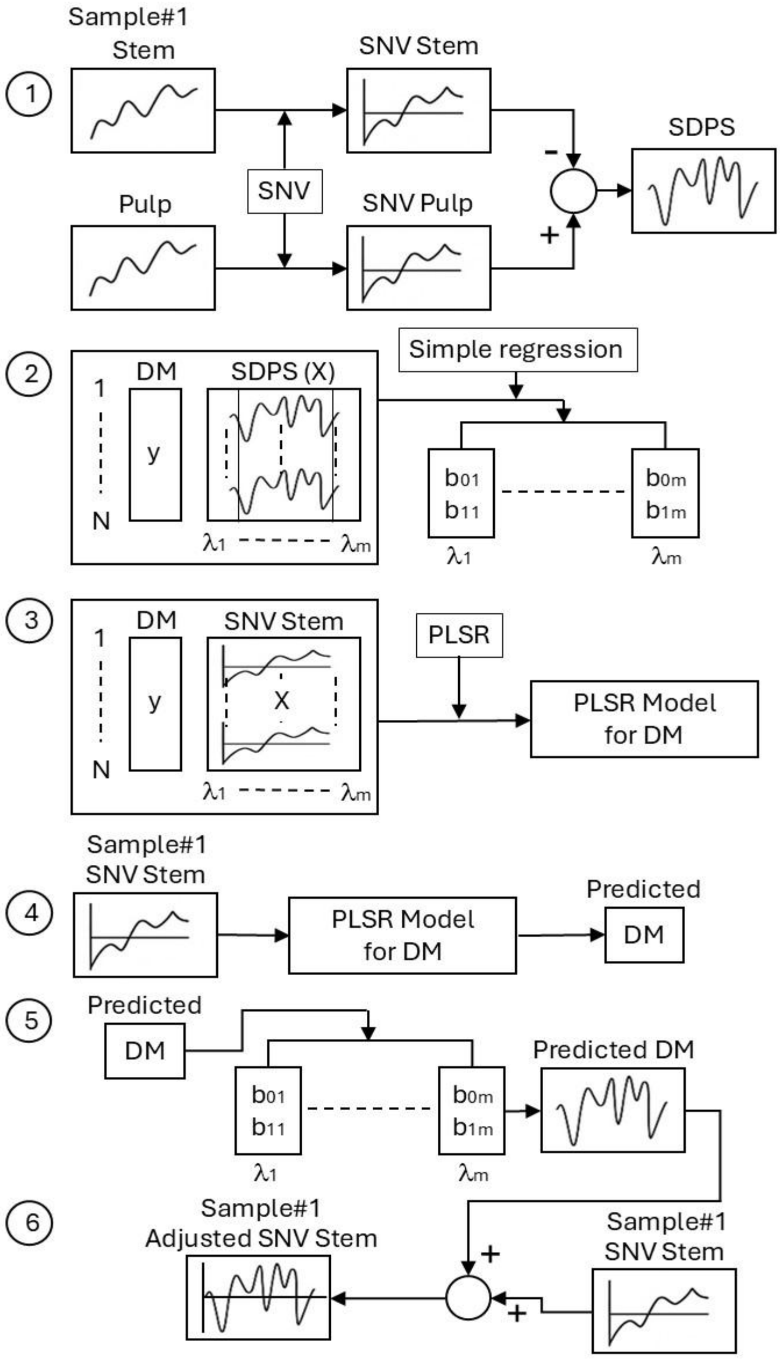

Figure 3. Diagram depicting steps for adjustment of SNV stem spectra using spectral difference (SDPS). SNV = Standard normal variate, DM = Dry matter).

2.4.3. Model development

Prior to model development, univariate outlier detection was performed using standardized DM values. Samples with standardized values greater than +3 or less than −3 were considered outliers and excluded from subsequent analyses [15]. Then, multivariate outlier detection was conducted using Mahalanobis distance analysis to identify spectra that deviated substantially from the overall spectral population [15]. Unlike standard Euclidean metrics, Mahalanobis distance measures the distance between a data point and the center of a distribution by accounting for variable correlations and scale differences. Spectral classification was then determined using probability values (p) derived from the chi-squared distribution function. Samples with $p < 0.05$ were identified as outliers. Based on these criteria, two outlying samples were identified in the test samples and removed from the testing set.

The durian samples were divided into a training set (108 samples for calibration) and a testing set (34 samples for prediction) at an approximate 3:1 ratio, ensuring similar variances in DM content between the two sets (Table 1). This was done by first arranging the samples in ascending order of DM values and then, within each group of four samples, assigning the first three to the training set and the last to the testing set.

All prepared spectra were used to develop prediction models using partial least squares regression in the MATLAB software (R2021a; MathWorks; Natick, MA, USA). During model calibration, spectral data and the DM content from the training set were used to build the prediction model. The optimal number of latent variables was determined based on leave-one-out cross-validation as the number yielding the minimum RMSECV. To avoid overfitting, no additional latent variables were included once the decrease in RMSECV became negligible. Then, the optimized model was used to predict the DM content of samples in the test set. The prediction performances were compared based on the correlation coefficient (R), coefficient of determination (R^2), standard error of prediction (SEP), root mean square error of prediction (RMSEP), bias, and ratio of standard deviation of the reference data to SEP of the prediction set (RPD).

Table 1. Descriptive statistics of dry matter content (%) in durian pulp of training and test sets.

	Number of samples	Minimum	Mean	Maximum	Standard deviation
Training set (Calibration)	108	24.73	35.74	40.26	2.85
Test set (Prediction)	34	26.93	35.95	40.25	2.78

2.5. Development of classification model

Classification models were developed for binary and three-class maturity classification. Initially, for the binary classification, the 32% DM threshold specified by the National Bureau of Agricultural Commodity and Food Standards [6] was considered. However, applying this threshold resulted in a highly unbalanced dataset, with only 14 immature samples compared with 128 mature samples. To improve class balance and provide sufficient samples for model training and evaluation, the threshold was increased to 34%, resulting in 34 immature samples.

The adoption of a 34% DM threshold also has practical relevance. In commercial maturity assessments, misclassifying an immature fruit as mature (false negative) is generally more critical than misclassifying a mature fruit with DM values between 32% and 34% as immature (false positive) [16].

The former may lead to harvesting fruit before adequate maturity is reached, whereas the latter results in only a more conservative harvest decision. Accordingly, a 34% DM threshold was adopted for binary classification to enhance the detection of immature fruit while achieving a more balanced class distribution. Samples with DM ≤ 34% were classified as immature, whereas those with DM > 34% were grouped into the mature/over-mature class.

For the three-class classification, samples were categorized as immature (DM ≤ 34%), mature (34% < DM < 37%), and overmature (DM ≥ 37%). The over-mature category corresponded to the stage associated with maximum eating quality [17]. Samples within each class were divided randomly into training and test sets using the same procedure employed for the regression model development.

For comparison, two models that achieved the best prediction performance, one based on unprocessed- or preprocessed stem spectra and the other using adjusted SNV stem spectra, were employed to develop a machine learning-based classification model. SVM has been identified as an optimal method for durian quality prediction [12, 18]; hence, it was used in this study. The classification model was built using the Classification Learner Toolbox in MATLAB (R2021a; Mathworks; Natick, MA, USA). SVM is a supervised algorithm that operates by finding the optimal hyperplane to separate data points within a continuous feature space [19]. This is achieved by maximizing the margin (i.e., the distance to the nearest data points, known as support vectors) with the primary objective of minimizing classification errors, especially for points outside the margin of tolerance (epsilon).

The NIR spectral data of the stem in the training set were utilized to develop the models. To mitigate overfitting, a K-fold cross-validation strategy was applied, with 5-fold cross-validation used as the default setting in the Classification Learner Toolbox. The performance of a classification model can be evaluated using metrics, such as accuracy, recall, and precision, as defined in equations (5)–(7), respectively:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

These performance metrics were calculated from the four standard confusion matrix elements: True positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), with the immature durian class defined as the positive class.

Accuracy indicates the proportion of correctly classified instances out of the total number of predictions. It ranges from 0 to 1 or can be represented as a percentage. Precision focuses on the positive predictions, quantifying the proportion of correctly identified positive cases out of all instances predicted as positive. Like accuracy, it is expressed between 0 and 1 or as a percentage. Recall, also known as sensitivity, measures the model's ability to capture all actual positive instances in the dataset. It, too, is reported as a value between 0 and 1 or as a percentage [20].

3. Results and discussion

3.1. NIR spectral characteristics with variation in maturity

The effect of maturity on stem NIR absorbance was examined based on generating three spectra by averaging the measurements from the immature, mature, and overmature sample groups, with averaged DM contents of 32.5%, 36.4%, and 38.5%, respectively (Figure 4). The NIR absorbance of the stem decreased as the DM content increased, indicating that stems of more mature durian reflected a greater amount of NIR light. In general, the spectra were characterized by strong water absorption bands at around 1940 nm and 1450 nm, associated with the first overtone of the O–H stretching vibration and a combination of O–H stretching and bending vibrations [21]. A weaker water absorption band occurred near 1200 nm, generally due to the first overtone of O–H stretching, the second overtone of C–H stretching, and a combination band of N–H stretching and bending [20]. Additionally, there were weak absorption features around 1800 nm, associated with C–H combination bands involving C–H stretching and bending vibrations [22].

Subsequently, the absorbance spectra were preprocessed using the Savitzky-Golay second derivative algorithm to improve spectral resolution and facilitate detailed band interpretation (Figure 5). There were three distinct peaks at 1150 nm, 1400 nm, and 1800 nm, for immature, mature, and overmature sample groups, respectively, each showing a monotonic trend with increasing maturity. The peak at 1150 nm corresponded to C–H second overtones and combination bands involving O–H [22]. The 1400 nm wavelength corresponded primarily to O–H bond overtones, specifically a combination of the first overtone and stretch vibrations of the O–H bond, often found in water (H_2O) [21]. The absorption band associated with C–H combination vibrations at 1800 nm, which appeared in the original absorbance spectra in Figure 4, became more clearly resolved in Figure 5 as a result of the second derivative preprocessing. The absorption band around 1800 nm has been associated with lignin. For example, Tsuchikawa and Siesler [23] reported that the 1780–1820 nm region included contributions from lignin based on analysis using chemometric modeling. Chattavongsin and Siriphanich [24] investigated the anatomical changes in durian stems at different stages of maturity. They observed that in Monthong durians, the number of cortical sclereids in the outer region of the lower stem increased with maturation. Typically, these sclereids develop a secondary cell wall that is often lignified, rendering them highly rigid and impermeable to water [25].

The spectral difference, obtained by subtracting the SNV stem spectra from the SNV pulp spectra, is shown in Figure 6. The adjusted SNV stem spectra of samples with varying DM contents are presented in Figure 7. When compared with the second derivative spectra (Figure 5), it is noteworthy that the adjusted stem spectra exhibited a relationship with DM across all dominant peaks, whereas such a relationship was observed only at certain peaks in the second derivative spectra. These findings indicated that the adjustment could potentially strengthen the correlation between the stem spectra and DM values.

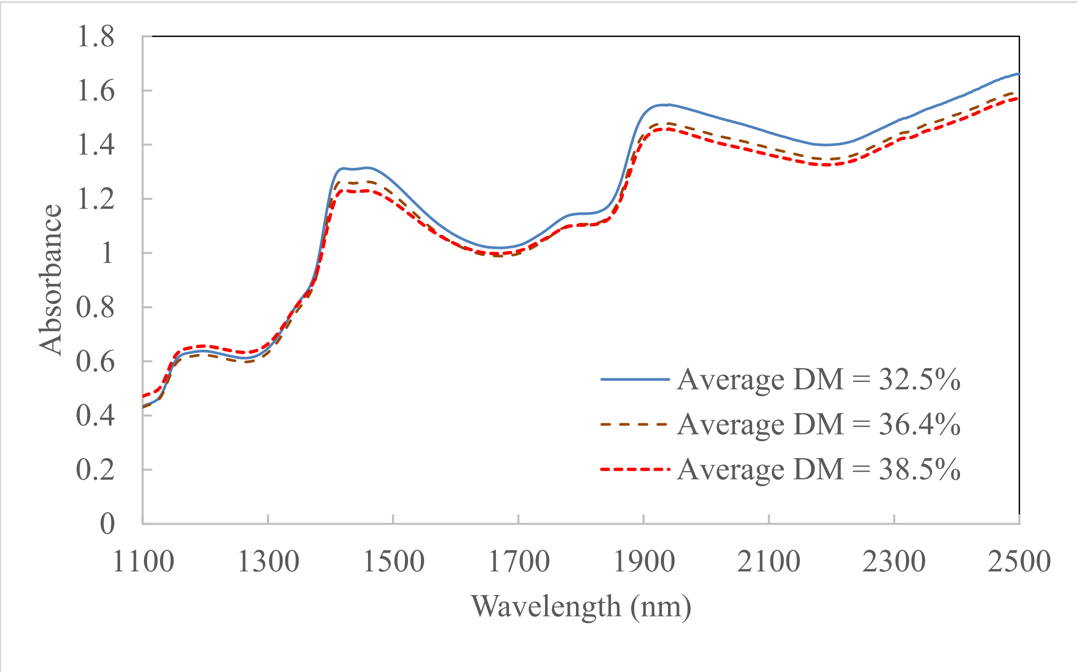

Figure 4. Effect of dry matter content of pulp on the near-infrared absorption spectra of durian stem.

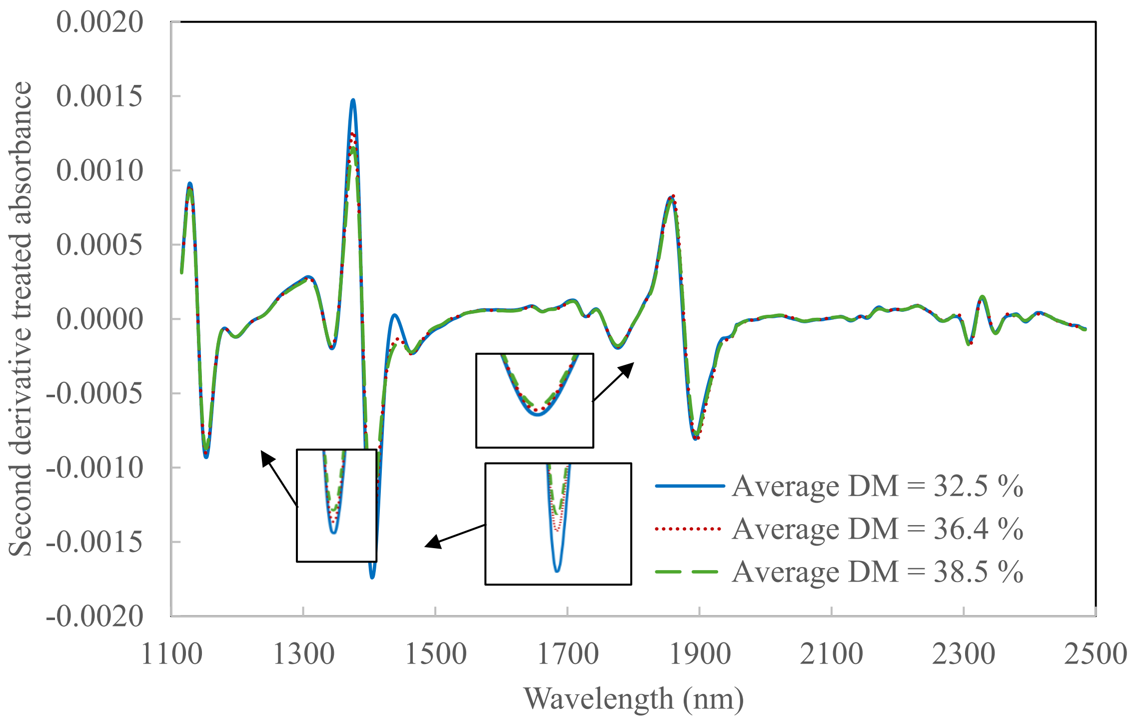

Figure 5. Variation in second derivative treated absorbance of a durian stem as affected by dry matter content of pulp.

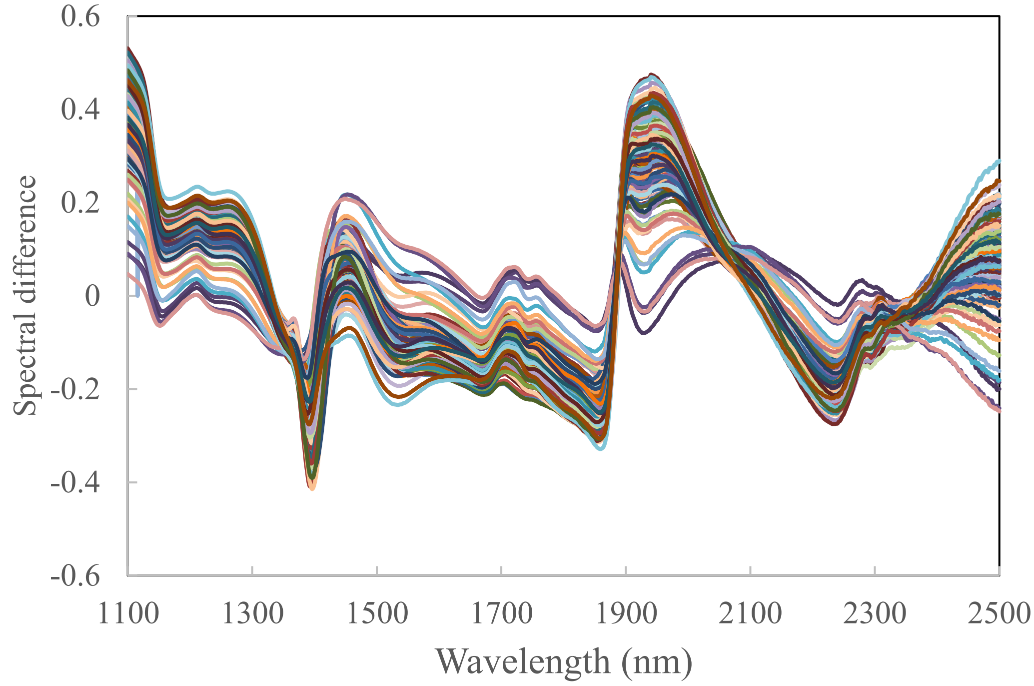

Figure 6. Spectral difference created by subtracting the SNV stem spectra from the SNV pulp spectra.

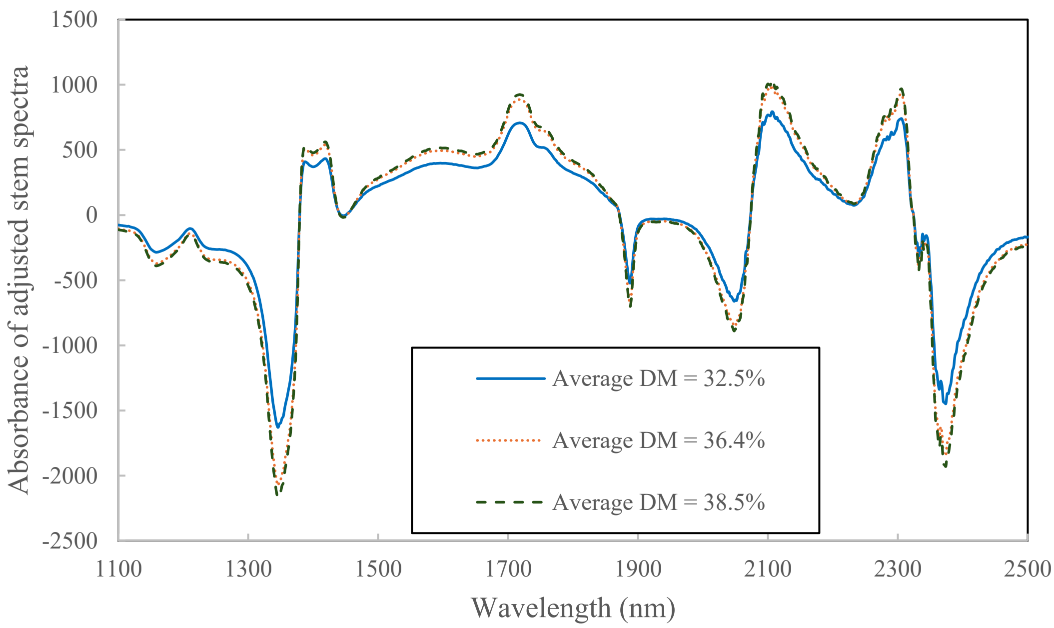

Figure 7. Effect of dry matter content of pulp on adjusted stem spectra.

3.2. Prediction performance of DM based on stem NIR spectra

The performance of the prediction models based on the stem NIR spectra with unprocessed and various preprocessing methods is summarized in Table 2. Among the conventional preprocessing techniques, the SNV-D1 model provided the best overall performance, achieving an Rp of 0.761, an RMSEP of 1.924%, and an RPD of 1.28. The combination of scatter correction and the first derivative transformation likely enhanced the spectral features associated with pulp DM content while reducing background variation. In contrast, the models developed using unprocessed spectra, SNV alone, D1 alone, and D2 alone showed lower predictive ability, indicating that these preprocessing methods were less effective in capturing the relationship between stem spectra and pulp DM content.

Table 2. Prediction performance of partial least squares regression models developed using different spectral preprocessing methods and adjusted stem spectra.

	None	SNV	D1	D2	SNV-D1	SNV-D2	Adjusted SNV stem spectra
Number of factors	13	9	7	11	11	11	3
R_c	0.827	0.771	0.758	0.893	0.913	0.857	0.922
R_c^2	0.684	0.594	0.575	0.797	0.833	0.734	0.848
RMSEC (%)	1.698	1.925	1.969	1.362	1.233	1.557	1.178
SEC (%)	1.706	1.934	1.978	1.368	1.239	1.565	1.178
Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R_{cv}	0.667	0.637	0.654	0.659	0.649	0.580	0.773
R_{cv}^2	0.436	0.396	0.431	0.402	0.363	0.230	0.600
RMSECV (%)	2.290	2.369	2.300	2.358	2.433	2.675	1.928
SECV (%)	2.300	2.380	2.310	2.358	2.445	2.687	1.937
Bias	−0.031	−0.001	0.022	0.037	0.025	−0.067	−0.001
R_p	0.608	0.444	0.584	0.724	0.761	0.731	0.828
R_p^2	0.270	0.197	0.256	0.475	0.508	0.482	0.647
RMSEP (%)	2.375	2.865	2.397	2.014	1.924	1.973	1.630
SEP (%)	2.407	2.899	2.431	2.002	1.872	1.959	1.621
Bias	0.076	0.198	0.008	−0.401	−0.547	−0.410	−0.324
Slope	0.560	0.422	0.511	0.649	0.714	0.809	0.842
Intercept	15.884	20.982	17.583	12.235	9.797	7.241	5.492
RPD	0.99	0.82	0.98	1.19	1.28	1.22	1.47

The adjusted SNV stem spectra model yielded the highest prediction performance among the evaluated models, with an Rp of 0.828, an RMSEP of 1.630%, and an RPD of 1.47 (Figure 8). Although the improvement over the best conventional model (SNV-D1) was not large, the adjusted spectra consistently provided slightly better results in calibration, cross-validation, and independent prediction. For example, Rp increased from 0.761 to 0.828, while RMSEP decreased from 1.924% to 1.630%. Based on these results, the spectral adjustment procedure may have captured additional information related to pulp DM content beyond that contained in the conventionally preprocessed stem spectra.

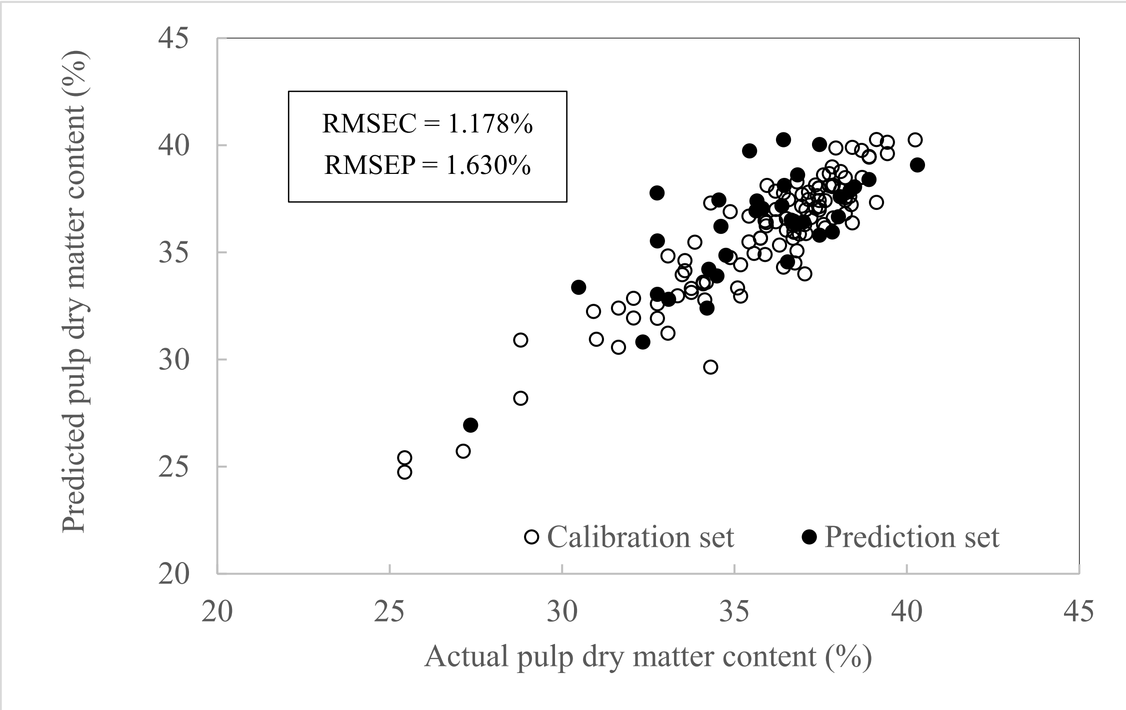

Figure 8. Scatter plot of actual versus predicted dry matter content based on the model developed using adjusted NIR stem spectra based on the spectral difference between stem and pulp spectra.

A notable feature of the adjusted SNV model was its simpler structure, requiring only three latent variables compared with 7–13 latent variables for the unprocessed and preprocessed models. Despite the reduced model complexity, the adjusted spectra achieved the highest values for Rc (0.922) and Rcv (0.773), together with the lowest RMSEC (1.178%) and RMSECV (1.928%) values. The adjusted SNV model also had a smaller calibration-prediction error gap (RMSEC = 1.178% and RMSEP = 1.630%; difference ≈ 38%) than the unprocessed and preprocessed models, which generally had larger discrepancies between the calibration and prediction errors. Although the difference remained above the 20% threshold recommended for highly robust models [26], the adjusted SNV model demonstrated comparatively better robustness and generalization to unseen samples.

Overall, based on these results, spectral adjustment slightly improved the indirect prediction of durian pulp DM content from the stem spectra. Although the enhancement in model performance was not substantial, the approach offered an alternative strategy for extracting additional relevant information from stem spectra.

Note: Rc^2 = Coefficient of determination for calibration, Rc = Correlation coefficient of calibration, Rcv^2 = Coefficient of determination for cross-validation, Rcv = Correlation coefficient of cross-validation, Rp^2 = Coefficient of determination for prediction, Rp = Correlation coefficient of prediction, RMSEC = Root mean square error of calibration, SEC = Standard error of calibration, RMSECV = Root mean square error of cross-validation, SECV = Standard error of cross-validation, RMSEP = Root mean square error of prediction, SEP = Standard error of prediction, RPD = Ratio of standard deviation of the reference data to SEP of prediction set, SNV = Standard normal variate, D1

= First derivative, D2 = Second derivative.

Variable Importance in Projection (VIP) analysis was performed on the adjusted spectra model to identify the wavelengths that contributed most to the prediction of pulp DM content. VIP is a metric derived from partial least squares regression that quantifies the overall contribution of each predictor variable to the model by considering the amount of variance explained in the predictor matrix and its contribution to explaining the response variable. In general, wavelengths with VIP values greater than 1 are regarded as important variables in the PLS model because they contribute more than the average variable to explaining the response variable [27].

The VIP analysis identified major peaks at 1352 and 2372 nm, indicating that these wavelengths contributed mainly to the prediction of pulp dry matter content. The wavelength at 1352 nm corresponds primarily to the first overtone of O–H vibrations associated with water absorption [28]. It was also near a spectral region commonly attributed to the second overtone of C–H stretching vibrations and combination bands involving C–H deformation vibrations, which are characteristic of carbohydrates and lipids [29]. The VIP peak at 2372 nm fell within a strong absorption region associated with the O–H second overtone [30].

Overall, the VIP-selected wavelengths corresponded to chemically meaningful absorption bands associated with O–H and C–H functional groups. Their importance in the model suggested that the prediction of pulp DM content was influenced by spectral information related to moisture status and the accumulation of carbohydrates and other dry matter constituents.

3.3. Performance of maturity classification based on adjusted SNV stem spectra

3.3.1. Three-class classification

Table 3 and Table 4 present the three-class classification performance of the SVM models developed from the SNV-D1 and the adjusted SNV stem spectra for distinguishing durian maturity stages of immature, mature, and over-mature. Model evaluation was based on accuracy, precision, and recall for the training and test datasets.

The SNV-D1 model achieved moderate training performance (accuracy = 0.657, precision = 0.682, recall = 0.651). Although the immature class showed high precision (0.800), its recall was limited (0.615), indicating frequent misclassification of true immature samples. The mature and over-mature classes exhibited lower precision, primarily due to overlapping spectral patterns. For the test set, accuracy was slightly lower (0.647), with good recall for the immature fruit class (0.875), indicating that the model minimized false negatives for this class, which is a critical attribute for preventing consumer dissatisfaction.

The adjusted SNV model demonstrated comparably balanced results. Training accuracy, precision, and recall were 0.657, 0.670, and 0.648, respectively. For the test set, the average accuracy dropped slightly to 0.647, with macro precision improving to 0.711, while macro recall was 0.633. Overall, the adjusted SNV model performed similarly to the SNV-D1 model with a marginal difference in class-specific trade-offs.

Table 3. Confusion matrices for three-class maturity classification using support vector machine (SVM) models based on SNV-D1 stem spectra and adjusted SNV stem spectra.

SVM model based on SNV-D1 stem spectra							
Training set				Test set (prediction)			
Actual maturity	Predicted maturity			Actual maturity	Predicted maturity		
	Immature	Mature	Overmature		Immature	Mature	Overmature
Immature	16	7	3	Immature	7	1	0
Mature	3	24	11	Mature	4	5	3
Overmature	1	12	31	Overmature	0	4	10
SVM model based on adjusted SNV stem spectra							
Training set				Test set (prediction)			
Actual maturity	Predicted maturity			Actual maturity	Predicted maturity		
	Immature	Mature	Overmature		Immature	Mature	Overmature
Immature	16	8	2	Immature	5	3	0
Mature	4	21	14	Mature	0	5	7
Overmature	1	8	34	Overmature	0	2	12

Table 4. Comparison of classification performance for three maturity classes between models developed using SNV-D1 stem spectra and adjusted SNV stem spectra.

SVM model based on SNV-D1 stem spectra					
Dataset	Metric	Maturity class			Average
		Immature	Mature	Overmature	
Training	Accuracy				0.657
	Precision	0.800	0.558	0.689	0.682
	Recall	0.615	0.632	0.705	0.651
Test	Accuracy				0.647
	Precision	0.636	0.500	0.769	0.635
	Recall	0.875	0.417	0.714	0.669
SVM model based on adjusted SNV stem spectra					
Dataset	Metric	Maturity class			Average
		Immature	Mature	Overmature	
Training	Accuracy				0.657
	Precision	0.762	0.568	0.680	0.670
	Recall	0.615	0.538	0.791	0.648
Test	Accuracy				0.647
	Precision	1.000	0.500	0.632	0.711
	Recall	0.625	0.417	0.857	0.633

3.3.2. Two-class classification

In commercial practice, durian fruits are classified into immature and mature categories according to the standards of National Bureau of Agricultural Commodity and Food Standards [6]. Accordingly, in further analysis, we focused on the binary classification of durian samples based on the DM content threshold of 34%. When sorting durians for export, removing immature fruits is crucial because false negatives (immature fruits classified as mature) may result in immature fruit entering the export chain. Consequently, precision and recall were suitable metrics for evaluating the classification performance [31]. Precision measures the accuracy of identifying immature fruits among those predicted as such, while recall indicates the proportion of actual immature fruits that were correctly identified. However, false negatives are critical in this classification task, as they indicate that immature samples are incorrectly classified as mature. Therefore, recall, which measures the proportion of correctly identified immature samples out of all actual immature samples, is a more appropriate performance metric in this context.

The two-class classification performance of the SVM models developed from the SNV-D1 and the adjusted SNV stem spectra is summarized in Tables 5 and 6. Based on the classification results, the SNV-D1 and adjusted SNV stem spectra models achieved comparable performance in the training and test sets. The overall accuracy increased only slightly from 0.880 to 0.889 in the training set and from 0.853 to 0.882 in the test set following the spectral adjustment, indicating that the adjustment provided a marginal improvement in classification performance.

Table 5. Confusion matrices for two-class maturity classification using support vector machine (SVM) models based on SNV-D1 stem spectra and adjusted SNV stem spectra.

SVM model based on SNV-D1 stem spectra					
Training set			Test set (prediction)		
Actual maturity	Predicted maturity		Actual maturity	Predicted maturity	
	Immature	Mature		Immature	Mature
Immature	15 (TP)	11 (FN)	Immature	8	0
Mature	2 (FP)	80 (TN)	Mature	5	21
SVM model based on adjusted SNV stem spectra					
Training set			Test set (prediction)		
Actual maturity	Predicted maturity		Actual maturity	Predicted maturity	
	Immature	Mature		Immature	Mature
Immature	19	7	Immature	8	0
Mature	5	77	Mature	4	22

Note: TP = True Positive, FN = False Negative, FP = False Positive, TN = True Negative.

With the training set, the adjusted SNV model improved recall of the immature class from 0.577 to 0.731, while recall of the mature class decreased slightly from 0.976 to 0.939. Consequently, the average recall increased from 0.777 to 0.835. Similarly, the average precision changed only modestly, from 0.881 for the SNV-D1 model to 0.855 for the adjusted SNV model. Based on these results, the spectral adjustment mainly redistributed the classification performance between the two classes rather than producing a substantial overall improvement.

In the test set, both models achieved perfect recall (1.000) for the immature class, indicating that all immature fruits were correctly identified. The adjusted SNV model produced a slight improvement in mature-class recall (from 0.808 to 0.846), which contributed to a modest increase in the average recall (from 0.904 to 0.923). Likewise, the overall accuracy increased by only 2.9 percentage points (from 0.853 to 0.882). In addition, the average precision improved slightly (from 0.808 to 0.833).

Overall, the adjusted SNV model consistently outperformed the SNV-D1 model across most evaluation metrics for the two-class classification; however, the magnitude of the improvement was relatively small. The adjusted SNV approach provided a modest enhancement in classification robustness, particularly through a slight reduction in the misclassification of mature fruits, while maintaining complete detection of immature fruits in the test set.

Table 6. Comparison of classification metrics for two maturity classes between models developed using SNV-D1 stem spectra and adjusted SNV stem spectra.

SVM model based on SNV-D1 stem spectra				
Dataset	Metric	Maturity class		Average
		Immature	Mature	
Training	Accuracy			0.880
	Precision	0.882	0.879	0.881
	Recall	0.577	0.976	0.777
Test	Accuracy			0.853
	Precision	0.615	1.000	0.808
	Recall	1.000	0.808	0.904
SVM model based on adjusted SNV stem spectra				
Dataset	Metric	Maturity class		Average
		Immature	Mature	
Training	Accuracy			0.889
	Precision	0.792	0.917	0.855
	Recall	0.731	0.939	0.835
Test	Accuracy			0.882
	Precision	0.667	1.000	0.833
	Recall	1.000	0.846	0.923

4. Conclusions

The NIR spectra acquired from the durian stems had potential for the indirect assessment of the pulp DM content and binary maturity classification. The proposed spectral adjustment technique, which incorporated pulp spectral information and DM values into stem spectra through the spectral difference between the pulp and the stem spectra, provided only modest improvements over conventional preprocessing methods in DM prediction and SVM-based maturity classification. While these findings suggested that spectral adjustment could be an alternative approach for enhancing stem spectral information, the performance gains were limited. Furthermore, this study should be regarded as a proof-of-concept study. Additional validation using independent samples collected across seasons, orchards, production lots, and measurement conditions is necessary to evaluate the robustness and generalizability of the proposed approach.

Although the measurements were conducted on the peeled surface of stems detached from the fruit due to instrument limitations, which prevented the method from being non-destructive, the experiment was performed at the proof-of-concept stage to establish the feasibility of the approach under controlled conditions. Further research is required to evaluate its performance on unpeeled stems attached to the fruit using instruments that support such measurements.

Author contributions

Kaewkarn Phuangsombut: Methodology, Data curation, Validation; Arthit Phuangsombut: Visualization, Investigation, Methodology; Sirinad Noypitak: Data curation, Resources, Software; Anupun Terdwongworakul: Writing—original draft, Writing—review & editing, Supervision, Conceptualization, Formal analysis. All authors have read and agreed to the published version of the manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors gratefully acknowledge research support from the Faculty of Engineering at Kamphaeng Saen, Kasetsart University, Thailand.

Conflict of interest

The authors declare no conflict of interest.

References

1. Maninang JS, Wongs-Aree C, Kanlayanarat S, et al. (2011) Influence of maturity and postharvest treatment on the volatile profile and physiological properties of the durian fruit. *Int Food Res J* 18: 1067–1075.
2. Department of Agricultural Extension (DOAE) (2026) Summary of the fruit production season situation report. Fruit Situation Reporting System. Available from: https://simplefruit.doae.go.th/dashboard/index.
3. Percival SS, Findley B (2007) What's in Your Tropical Fruit? FSHN 07-08/FS144, 9/2007. *EDIS* 2007: No.19. https://doi.org/10.32473/edis-fs144-2007
4. Pascua OC, Cantila MS (1992) Maturity indices of durian (*durio zibethinus murray*). *Philipp J Crop Sci* 17: 119–124.
5. Siriphanich J (2011) Durian (*Durio zibethinus* Merr.). In: Yahia EM (Ed.), *Postharvest biology and technology of tropical and subtropical fruits,* Woodhead Publishing, 80–114. https://doi.org/10.1533/9780857092885.80
6. National Bureau of Agricultural Commodity and Food Standards (2003) Thai Agricultural Standard TAS 3-2003: Durian. R Thai Gov Gaz 120(115D).

7. Onsawai P, Phetpan K, Khurnpoon L, et al. (2021) Evaluation of physiological properties and texture traits of durian pulp using near-infrared spectra of the pulp and intact fruit. *Measurement* 174: 108684. https://doi.org/10.1016/j.measurement.2020.108684
8. Pokhrel DR, Sirisomboon P, Khurnpoon L, et al. (2023) Comparing machine learning and PLSDA algorithms for durian pulp classification using inline NIR spectra. *Sensors* 23: 5327. https://doi.org/10.3390/s23115327
9. Puttipipatkajorn A, Terdwongworakul A, Puttipipatkajorn A, et al. (2023) Indirect prediction of dry matter in durian pulp with combined features using miniature NIR spectrophotometer. *IEEE Access* 11: 84810–84821. https://doi.org/10.1109/ACCESS.2023.3303020
10. Saenphon C, Ditcharoen S, Malai C, et al. (2023) Total soluble solids, dry matter content prediction and maturity stage classification of durian fruit using long-wavelength NIR reflectance. *J Food Compos Anal* 124: 105667. https://doi.org/10.1016/j.jfca.2023.105667
11. Posom J, Saenphon C, Ditcharoen S, et al. (2025) Deep neural networks (DNNs) chemical compositions estimation of fresh durian in-line via near infrared spectroscopy. *J Food Compos Anal* 142: 107410. https://doi.org/10.1016/j.jfca.2025.107410
12. Ditcharoen S, Sirisomboon P, Saengprachatanarug K, et al. (2023) Improving the non-destructive maturity classification model for durian fruit using near-infrared spectroscopy. *Artif Intell Agric* 7: 35–43. https://doi.org/10.1016/j.aiia.2023.02.002
13. Imai M, Sangsoy K, Blasco J, et al. (2025) Enhancing dry matter prediction in durian using peduncle-based NIR spectroscopy across maturation rates. *Postharvest Biol Technol* 230: 113773. https://doi.org/10.1016/j.postharvbio.2025.113773
14. Sangwanangkul P, Siriphanich J (2000) Growth and maturation of durian fruit cv. Monthong. *Thai J Agric Sci* 33: 75–82.
15. Meier KJ, Brudney JL, Bohte J (2009) Applied statistics for public and nonprofit administration. Cengage Learning, Wadsworth, Boston.
16. Goodwin CJ, Goodwin KA (2020) Research in psychology: methods and design. 9th ed., Wiley.
17. Yongyut N, Baopa P, Meetha S, et al. (2025) Fruit quality and antioxidant content in durian (*Durio zibethinus* Murr.) cv. 'Monthong' in different maturity stages. *Horticulturae* 11: 432. https://doi.org/10.3390/horticulturae11040432
18. Ali MM, Hashim N, Shahamshah MI (2021) Durian (*Durio zibethinus*) ripeness detection using thermal imaging with multivariate analysis. *Postharvest Biol Technol* 176: 111517. https://doi.org/10.1016/j.postharvbio.2021.111517
19. Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14: 199–222. https://doi.org/10.1023/B:ASTCO.0000035301.49549.88
20. Raschka S, Mirjalili V (2019) Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing Ltd., Birmingham.
21. Workman J, Weyer L (2007) Practical guide to interpretive near-infrared spectroscopy. CRC Press.
22. Ciurczak EW, Igne B, Workman J Jr, et al. (Eds.) (2021) Handbook of near-infrared analysis. CRC Press, Boca Raton, FL.
23. Tsuchikawa S, Siesler HW (2006) Near-infrared spectroscopic monitoring of the diffusion process of deuterium-labeled molecules in wood. In: Stokke DD, Groom LH (Eds.), *Characterization of the Cellulosic Cell Wall,* Blackwell Publishing, 123–137. https://doi.org/10.1002/9780470999714.ch10

24. Chattavongsin R, Siriphanich J (1987) Fruit-stem anatomy of durians at various harvesting stages [in Thai with English abstract]. In: Proceedings of the 26th Academic Conference, Kasetsart University, Bangkok, 405–412.
25. Beck CB (2010) An introduction to plant structure and development: Plant anatomy for the twenty-first century. Cambridge University Press, Cambridge, UK, 17–24. https://doi.org/10.1017/CBO9780511844683
26. Williams PC, Norris KH (2001) Near-infrared technology in the agricultural and food industries. 2nd ed., American Association of Cereal Chemists.
27. Wold S, Johansson E, Cocchi M (1993) PLS-partial least squares projections to latent structures. In: 3D QSAR in Drug Design: Theory, Methods and Applications, Elsevier.
28. Currà A, Gasbarrone R, Bonifazi G, et al. (2022) Near-infrared transflectance spectroscopy discriminates solutions containing two commercial formulations of botulinum toxin type A diluted at recommended volumes for clinical reconstitution. *Biosensors* 12: 216. https://doi.org/10.3390/bios12040216
29. Beć KB, Grabska J, Huck CW, (2020) Near-infrared spectroscopy in bio-applications. Molecules 25(12):2948. https://doi.org/10.3390/molecules25122948
30. Osborne B, Fearn T, Hindle PH (1993) Practical NIR spectroscopy with applications in food and beverage analysis, 2nd ed., Addison-Wesley Longman Ltd.: Harlow, UK, 227.
31. Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. *Inf Process Manag* 45: 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)