RAMReS Sciences des Structures et de la Matière Vol. 10, No. 1 (2026) 1 - 11
http://publication.lecames.org/index.php/mat *ISSN: 2630-1180*

Identification of twelve frequently processed wood species from Ivorian forests by carpenters using near-infrared spectroscopy

N'guessan J. L. Lepetit[1,*], Niamke B. Florence[1], Ouattara H. A. Aziz[1], Bley-Atse B. Appolinaire[1], Yao N. Jean-Claude[1], Nadine Amusant[2], Adima A. Augustin[1]

[1]*Laboratory of Industrial Processes and Synthesis of the Environment and New Energies, National Polytechnic Institute Félix Houphouët-Boigny, Yamoussoukro, Côte d'Ivoire*
[2]*French Agricultural Research Centre for International Development Department, UMR Ecofog, BP 701, 91310 Kourou Cedex, France*

Received: 26 October 2025 / Received in revised form: 20 December 2025 / Accepted: 12 March 2026

***Abstract*:**

This study demonstrates the efficacy of near-infrared (NIR) spectroscopy combined with chemometric techniques for identifying twelve tropical wood species commonly processed in Ivorian carpentry workshops. Principal Component Analysis (PCA) was employed to explore spectral variability, revealing distinct groupings corresponding to differences in chemical composition, wood density, and anatomical characteristics. Although some overlap was observed between certain species, three major trends could be distinguished, indicating meaningful structural or chemical similarities within each group. Subsequently, a Partial Least Squares Discriminant Analysis (PLS-DA) model was developed using 864 spectral samples. The model achieved a classification accuracy of 97.8% with raw spectra, which increased to 98.96% after applying the first derivative as a preprocessing step. This improvement confirmed the utility of spectral preprocessing for enhancing signal separation and reducing scattering effects. Several species were classified with perfect accuracy, demonstrating highly distinctive spectral signatures. Others showed improved discrimination after preprocessing, suggesting closer chemical similarities that were better resolved through data treatment. The strong overall performance confirms the robustness of this approach. The results affirm that NIR spectroscopy, when combined with appropriate chemometric tools, offers a reliable, rapid, and non-destructive method for tropical wood identification. This approach is particularly valuable in contexts requiring species-level discrimination, such as timber traceability, quality control, or combating illegal logging in biodiverse tropical regions.

***Keywords* :** NIR spectroscopy ; Classification ; Identification ; Chemometric tools.

* Corresponding author:
Email address: *louis.nguessan21@inphb.ci (J.L.L. N'guessan)*
https://doi.org/10.70974/mat10126001

1. Introduction

Forestry activities have significantly contributed to the decline of forest cover in Côte d'Ivoire. Numerous commercially valuable and ecologically important hardwood species have been heavily exploited, leading to alarming levels of depletion. Species such as Ivorian Mahogany (*Khaya ivorensis*), Iroko (*Milicia excelsa*), Fraké (*Terminalia superba*), Badi (*Nauclea diderrichii*), Tiama (*Entandrophragma angolense*), Dabema (*Piptadeniastrum africanum*), Niangon (*Tarrietia utilis*), and Movingui (*Distemonanthus benthamianus*) are now considered at risk of extinction due to unsustainable logging practices [1]. This logging activity, whether legal or illegal, generates substantial amounts of wood waste, primarily in the form of sawdust, which pollutes urban areas and poses significant environmental challenges.

Sawdust, having no immediate value for industries or woodworking shops, is often discarded as waste, burned, or buried, thereby exacerbating environmental issues. However, this wood waste contains chemical compounds (primary and secondary metabolites) that could be valorized in agriculture, pharmaceuticals, wood preservation, and other applications.

The B4AP project (Bioactives for Agriculture and wood Preservatives), led by INP-HB (Institut National Polytechnique Félix HOUPHOUËT-BOIGNY), aimed to valorize the chemical compounds found in wood waste. However, the project identified constraints related to sawdust collected from carpentry workshops. Specifically, workshop sawdust is often a complex mixture of sapwood and heartwood from multiple species, making it impossible to valorize compounds from individual wood species. Another major constraint was the accurate identification of sawdust by species.

To address this issue, several major woodworking shops in Côte d'Ivoire (Yamoussoukro) were trained to collect and store sawdust separately by wood species. This approach resolved the problem of mixed sawdust. However, for accurate identification, a more precise method was essential, as some wood species, due to similar characteristics (visual, mechanical, chemical, morphological, etc.), could be confused by carpenters. For example, Dabema (*Piptadeniastrum africanum*) sawdust could be mistaken for Samba (*Triplochiton scleroxylon*) sawdust.

Near-infrared spectroscopy (NIRS) has proven to be a highly accurate technique for identifying sawdust by wood species [2]. This method enables the development of extensive spectral databases, thereby facilitating and accelerating species identification [3–5]. NIRS operates based on the interaction of electromagnetic radiation with molecular bonds such as C–H, S–H, N–H, and O–H present in organic materials [5]. To extract meaningful insights from spectral data, chemometric techniques are employed to correlate NIR spectra with the qualitative and quantitative properties of wood. These methods support the development of classification models, the design and optimization of experiments, and the calibration of multivariate regression models based on the material's spectral signature [6, 7].

Recent studies have highlighted the high potential of near-infrared spectroscopy (NIRS) combined with machine learning techniques for accurate wood species identification. For instance, Pan *et al.* [7] developed a multimodal fusion framework based on deep learning, integrating NIR spectral data with RGB imagery. This approach achieved identification accuracies exceeding 99% for complex wood samples. Similarly, other studies have demonstrated that combining hyperspectral imaging with advanced machine learning algorithms significantly enhances both the accuracy and efficiency of wood species classification [8, 9].

The sawdust species studied include Teak (*Tectona grandis* Linn), Fraké (*Terminalia superba*), Cedrela (*Cedrela odorata*), Hévéa (*Hevea brasiliensis*), Colatier (*Cola acuminata*), Bete (*Mansonia altissima*), Acajou (*Khaya ivorensis*), Dabema (*Piptadeniastrum africanum*), Acacia (*Acacia mangium*), Samba (*Triplochiton scleroxylon*), Gmelina (*Gmelina arborea*), and Iroko (*Milicia excelsa*). These species were selected based on a survey conducted among woodworking shops regarding the wood types they frequently receive and process.

The aim of this study is to develop accurate identification models using near-infrared spectroscopy combined with chemometric tools to correctly identify and classify various sawdust samples collected by trained workshops. Proper identification would promote more targeted valorization of the chemical compounds derived individually from each studied species.

2. Materials and methods

Twelve different tree species were selected, with 72 sawdust samples collected per species (Table 1). The sawdust (Figure 1) was ground separately by species using a Retsch SM100 grinder (Retsch® Group, Haan, Germany) fitted with a 0.5 mm mesh sieve. This grinding process generated uniform sawdust particles suitable for near-infrared (NIR) spectral acquisition. Following grinding, the samples were conditioned in a climate-controlled chamber maintained at $30 \pm 2°C$ and $80 \pm 5\%$ relative humidity for two weeks to stabilize their moisture content. The average moisture level for each species was then determined in triplicate. The samples obtained from carpenters consisted of sapwood–heartwood mixtures. Particle size distribution was carefully checked to minimize its influence. Three trees per species were used, all originating from the same forest exploitation area in the Yamoussoukro department. Regarding age, all trees were mature; however, since maturity age varies among species, precise tree ages were not available.

2.1. Near-infrared spectroscopy (NIR)

Spectral data were collected in diffuse reflectance mode using a portable microNIR Pro V3.0 spectrometer (900–1700 nm) developed by Viavi Solutions (Milpitas, CA, USA). Each measurement was performed directly on the surface of the wood sawdust, with thirty-two scans acquired per sample (Figure 2). The average spec-

trum was then computed for each. All measurements took place in an air-conditioned room maintained at approximately $20 \pm 2°C$ with a relative humidity of $65 \pm 5\%$. The analysis included a comparison between raw spectral data and preprocessed spectra. Preprocessing involved applying first and second derivatives (second-order polynomial) using a 15-point window on each side, followed by Standard Normal Variate (SNV) normalization.

2.2. Multivariate statistical analysis

Multivariate exploratory analyses and model calibration were performed using Chemoface software (version 1.66). Chemometric techniques (preprocessing, PCA, PLS-R, PLS-DA) were applied to facilitate species identification and discrimination by transforming chemical spectral data into mathematical models. These multivariate analyses were carried out on both raw and preprocessed data to assess the influence of preprocessing steps on model performance.

2.3. Discriminant analysis by partial least squares regression (PLS-DA)

In the PLS-DA framework, each wood species was treated as a categorical variable with distinct characteristics. Samples were classified into twelve categories, with, for example, category Ac representing *Khaya ivorensis* (Table 1). Binary values were then assigned: a value of 1 if a sample belonged to a given category, and 0 otherwise. Preliminary PLS-R models were developed to estimate membership for each of the twelve categories (Ac, Da, A, Sa, Gm, Ir, Te, Fr, Ce, He, Co, Be). These models were evaluated using the cross-validation coefficient (R^2_{cv}). For classification purposes, a sample was assigned to the category corresponding to the model yielding the highest predicted value. The performance of the PLS-DA models was further evaluated based on the classification accuracy observed in the graphical score plots. A total of 864 samples were used for the development and evaluation of the chemometric models.

Table 1
List of wood species studied.

No.	Code	Common name	Scientific name	Family
1	Ac	Acajou	*Khaya ivorensis*	Meliaceae
2	Da	Dabema	*Piptadeniastrum africanum*	Fabaceae
3	A	Acacia	*Acacia mangium*	Fabaceae
4	Sa	Samba	*Triplochiton scleroxylon*	Malvaceae
5	Gm	Gmelina	*Gmelina arborea*	Verbenaceae
6	Ir	Iroko	*Milicia excelsa*	Moraceae
7	Te	Teck	*Tectona grandis Linn*	Lamiaceae
8	Fr	Fraké	*Terminalia superba*	Combretaceae
9	Ce	Cedrela	*Cedrela odorata*	Meliaceae
10	He	Hévéa	*Hevea brasiliensis*	Euphorbiaceae
11	Co	Colatier	*Cola acuminata*	Sterculiaceae
12	Be	Bete	*Mansonia altissima*	Malvaceae

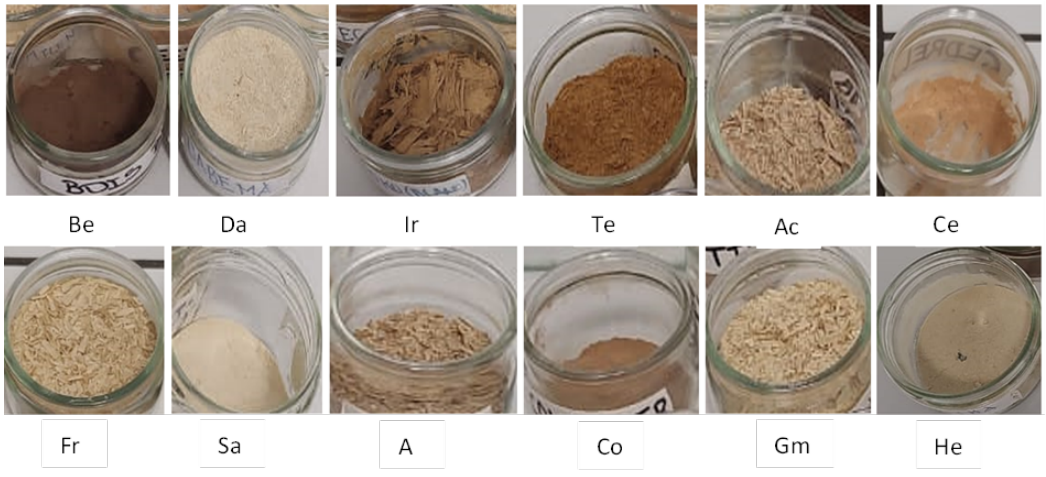

Fig. 1. Wood sawdust from various types of wood.

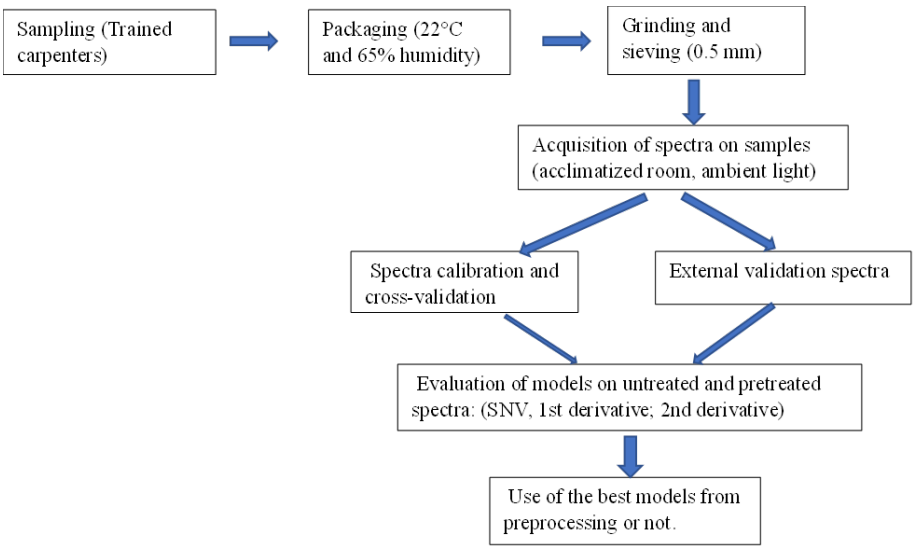

Fig. 2. Explicit methodology diagram.

The dataset was divided into two subsets: a calibration set subjected to leave-one-sample-out cross-validation, and an external validation set used for independent prediction. For calibration, 75% of the samples (648 samples) were selected using a stratified random sampling approach to ensure adequate representation of the different classes and the full range of variability. This subset was used to build the initial PLS models.

Internal performance was assessed using leave-one-sample-out (LOSO) cross-validation applied to the calibration set. At each iteration, a single sample was removed from the calibration set, the model was recalibrated on the remaining 647 samples, and the excluded sample was then predicted. This procedure was repeated 648 times, so that every calibration sample served once as the validation instance. This approach provides a robust and nearly unbiased estimate of internal performance (RMSECV, R^2_{CV}), as each validation step relies on an isolated sample. Finally, 25% of the samples (216 samples) were reserved for external validation. These samples were selected randomly and were never used during model construction. They were employed to evaluate the true predictive ability of the final model using standard performance indicators such as RMSEP and R^2_P.

3. Results

3.1. Material identification

For the identification of sawdust originating from these same wood species, a total of 864 spectra were collected from various sawdust samples. The collected spectra were used for model calibration and validation. For the calibration set, 75% of the spectra were used, and 25% of the spectra were used for validation.

3.2. Principal Component Analysis (PCA)

For the spectra collected from the sawdust samples, Principal Component Analysis (PCA) revealed that the raw data accounted for the largest portion of the total variance in the analysis space (Table 2). Preprocessing the dataset using the first derivative showed values similar to those of the untreated components. The dataset preprocessed with Standard Normal Variate (SNV) demonstrated a low explanatory power for wood sawdust. The first principal component (PC1) alone explained more than 99% of the variance in the untreated spectral dataset. Using only the first two principal components, it was possible to generate score plots forming three main groups. The first group is represented by the scores of Bete (Be) and Colatier (Co) (at the bottom of the plot); the second group, closely related to the first, includes the scores of Acajou (Ac), Dabema (Da), Hévéa (He), Gmelina (Gm), Cedrela (Ce), and Acacia (A); finally, the third group consists of scores of Samba (Sa), Teck (Te), Iroko (Ir), and Fraké (Fr). This exploratory analysis of spectral behavior was highly meaningful, using only the first two principal components (Figure 3).

3.3. Modelling (PLS-R and PLS-DA)

The modeling of each species using PLS-R with and without spectral preprocessing enabled the description of several parameters, which are presented in the tables. The different PLS-R models were obtained through cross-validation and used to test the predictive ability of the PLS-DA models.

The Partial Least Squares Regression (PLS-R) models developed from raw NIR spectra and their first derivatives allowed for the evaluation of the predictive ability of the chemical properties of twelve wood species (Table 3). The models built from raw spectra showed generally high performance, with coefficients of determination for calibration (R^2_c) ranging from 0.84 for Acacia (A) to 0.96 for Acajou (Ac), and low root mean square errors of calibration ($RMSE_c$), ranging from 0.07 for Teck (Te) to 0.15 for Dabema (Da) and Iroko (Ir).

Cross-validation confirmed the robustness of the models, with R^2_{cv} values ranging from 0.78 (Acacia, A) to 0.96 (Acajou, Ac), and $RMSE_{cv}$ values generally close to the $RMSE_c$. External set predictions (R^2_p) showed some variability depending on the species, ranging from 0.75 (Colatier, Co) to 0.91 (Acajou, Ac) using

first derivative preprocessing, with $RMSE_p$ values between 0.32 for Teck (Te) and 0.39 for Acajou (Ac), Dabema (Da), Iroko (Ir), and Gmelina (Gm).

Comparison between raw spectra and first derivatives indicated that the effect of preprocessing is highly species dependent. For some species, such as Acajou (Ac), Gmelina (Gm), Fraké (Fr), and Bete (Be), applying the derivative maintained or even improved predictive performance, as reflected in high R^2_p values (≥ 0.86). Conversely, for other species, such as Colatier (Co), Hévéa (He), and Cedrela (Ce), a degradation in performance was observed after derivation, with a notable decrease in R^2_p and an increase in $RMSE_p$.

Overall, the best-performing models were those for Colatier (Co), Bete (Be), Teck (Te), and Hévéa (He) based on raw spectra ($R^2_p \geq 0.86$, $RMSE_p \leq 0.21$), while Acacia (A), Dabema (Da), and Iroko (Ir) showed more limited predictive capabilities ($R^2_p \leq 0.81$).

Regarding model complexity, the number of latent variables (LVs) used varied moderately across species, ranging from 11 to 14, with a slightly higher average for derivative spectra. This stability suggests a moderate model complexity appropriate to the chemical diversity of the studied wood species.

These results highlight both the potential of near-infrared spectroscopy (NIRS) for predicting the chemical properties of tropical woods, and the need for finely tuned spectral preprocessing and modeling parameters for each species to optimize predictive performance.

The classification table reveals excellent overall performance of the model applied to the NIR spectra, with a correct classification rate of 97.80% for raw spectra and 98.96% after applying the first derivative (Table 4). Certain species, such as Acajou (Ac), Samba (Sa), Gmelina (Gm), Iroko (Ir), Teck (Te), and Bete (Be) were identified without error, regardless of the spectral preprocessing applied, reflecting their strong and distinct spectral signatures.

To provide a more detailed view of the confusion matrix derived from cross-validation, Table 5 presents the class-wise metrics (TP, FN, FP, Recall, Precision, Specificity, F1) for each species.

The use of the first derivative improved the classification of several species, notably Dabema (Da), Fraké (Fr), and Cedrela (Ce), by reducing misclassification errors with species having similar spectral profiles. These results highlight the value of spectral preprocessing in optimizing the discrimination of tropical woods through spectroscopic analysis combined with supervised classification. The improvement in classification accuracy observed after applying the SNV–SG preprocessing (98.96%) compared with the raw spectra (97.80%) was evaluated using bootstrap resampling (B = 1000). The mean performance difference was +1.16%, with a 95% confidence interval, indicating a statistically significant improvement ($p < 0.01$). As shown in Fig. 4, the VIP score plots derived from PLS models reveal clear differences between spectra pretreated with the first derivative and those left unpretreated, underscoring the impact of preprocessing on the model outcomes.

In order to complement these findings, Table 6 reports the detailed classification metrics obtained from the confusion matrix using spectra treated with the first derivative.

Furthermore, Table 7 summarizes the external validation results, highlighting the predictive ability of the PLS-DA models when applied to independent datasets.

To illustrate the robustness of the approach, Table 8 provides the bootstrap resampling statistics, confirming the stability and reliability of the classification accuracy improvements.

Finally, Table 9 presents the variable importance in projection (VIP) scores, which identify the spectral regions most responsible for species discrimination.

Table 2
Explanation of the variation X (in %) across the values of the principal components for spectra from sawdust.

	Untreated	SNV	1st derivative	2nd derivative
PC 1	99.39	98.70	99.07	98.98
PC 2	99.95	99.85	99.65	99.44
PC 3	99.99	99.94	99.85	99.75
PC 4	99.99	99.97	99.91	99.89

*PC x (x = 1, 2, 3 and 4) refers to the cumulative explained variance.

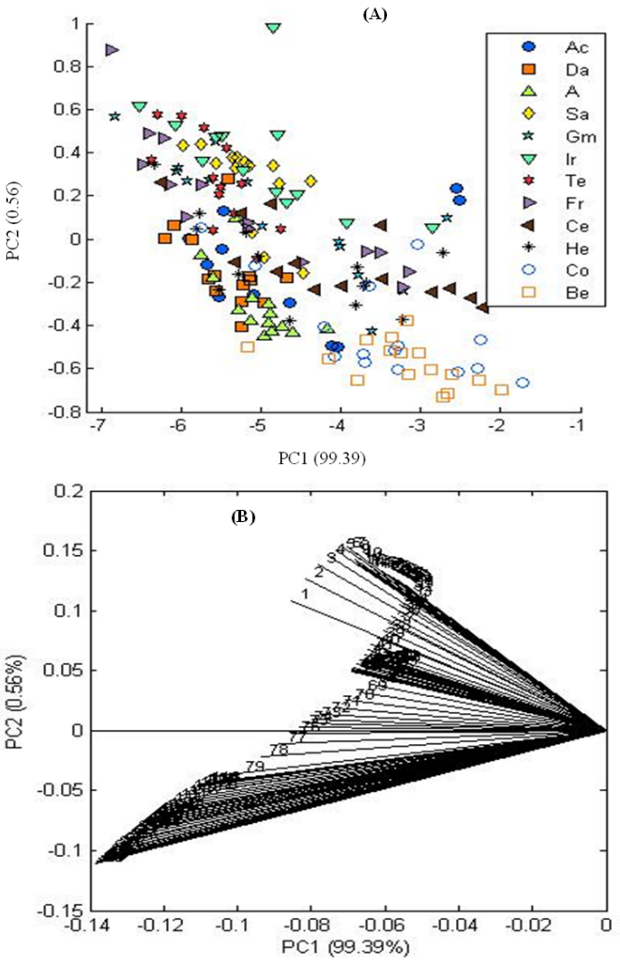

Fig. 3. (A) PCA results for wood sawdust and (B) loading plot.

Table 3
Statistics of PLS-R models used to classify PLS-DA based on spectra taken from wood sawdust.

Species	RMSEc	R_c^2	RMSEcv	R_{cv}^2	R_p^2	RMSEp	LV
			Untreated				
Ac	0.08	0.96	0.09	0.96	0.90	0.39	12
Da	0.15	0.88	0.17	0.85	0.81	0.39	13
A	0.14	0.84	0.17	0.78	0.77	0.33	12
Sa	0.11	0.90	0.12	0.89	0.83	0.33	13
Gm	0.11	0.89	0.13	0.89	0.87	0.17	13
Ir	0.15	0.88	0.17	0.85	0.81	0.39	11
Te	0.07	0.94	0.08	0.91	0.90	0.21	12
Fr	0.11	0.87	0.12	0.85	0.85	0.30	13
Ce	0.14	0.85	0.14	0.82	0.83	0.28	14
He	0.12	0.88	0.13	0.86	0.87	0.15	13
Co	0.11	0.90	0.13	0.88	0.88	0.12	11
Be	0.11	0.93	0.13	0.90	0.86	0.12	13
			1st Derivative				
Ac	0.08	0.96	0.09	0.95	0.91	0.39	13
Da	0.13	0.86	0.19	0.81	0.82	0.39	13
A	0.14	0.82	0.17	0.76	0.76	0.33	13
Sa	0.11	0.90	0.12	0.88	0.84	0.34	13
Gm	0.12	0.94	0.12	0.91	0.88	0.39	13
Ir	0.11	0.85	0.19	0.81	0.81	0.39	12
Te	0.10	0.95	0.12	0.93	0.89	0.32	11
Fr	0.11	0.87	0.12	0.87	0.86	0.35	13
Ce	0.13	0.83	0.13	0.80	0.77	0.33	13
He	0.13	0.84	0.14	0.80	0.78	0.37	12
Co	0.14	0.82	0.16	0.78	0.75	0.39	12
Be	0.11	0.94	0.12	0.91	0.88	0.39	12

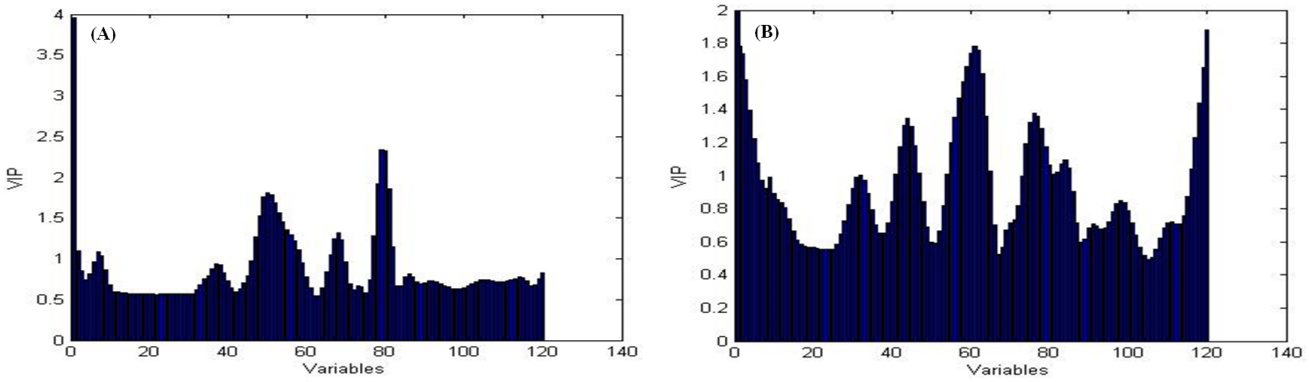

Fig. 4. VIP score plots derived from PLS models: (A) spectra pretreated with the first derivative; (B) spectra left unpretreated.

4. Discussion

4.1. Principal Component Analysis

Principal Component Analysis (PCA) scores may or may not reveal distinct groupings, as demonstrated in this study. In certain cases, such as in the work of Xue *et al.* [10] on *Guiboutia* species identification, PCA successfully produced clearly separated clusters, enabling effective discrimination at the species or family level in wood analysis. However, overlaps between data points, sometimes extensive, can occur, as also noted by Xue *et al.* [10]. A linear distribution of samples is also frequently observed in PCA plots, and in our study, this distribution could be separated into three distinct groups. The PCA of the spectral data provided insight into the chemical characteristics of both wood flour and whole wood samples, revealing patterns of clustering. These groupings are likely influenced by multiple factors, including wood density, chemical composition, and anatomical structure. Since near-infrared (NIR) spectroscopy reflects variations in chemical constituents such as cellulose, hemicellulose, lignin, and extractives, each wood species displays a unique spectral signature corresponding to its specific chemical profile. Thus, the observed groupings may reflect similarities in the chemical and/or physical prop-

erties of the wood samples [2]. Indeed, the groups at the low and high ends of the PCA are characterized by high-density woods such as Bete, Iroko, and teak. However, the intermediate group is more characterized by medium-density or low-density woods [11]. Indeed, in Côte d'Ivoire, the variability of timber properties is considerable due to the wide diversity of exploited species. Several studies on African tropical woods have shown that near-infrared (NIR) spectral signatures are sufficiently sensitive to capture interspecific differences. For instance, Mezui *et al.* [12] demonstrated robust differences in the NIR spectra of 25 tropical hardwood species from Gabon, reflecting chemical variations correlated with biological and ecological traits.

Near-infrared spectroscopy, combined with density information, therefore constitutes a particularly powerful tool for the differentiation and characterization of Ivorian woods. By simultaneously capturing variations in chemical composition (cellulose, hemicelluloses, lignin, and extractives) and anatomical structure that determine wood density, NIRS enables the effective discrimination of species, including those that are visually similar or commercially confused [12].

4.2. PLS-DA model calibration and validation

The results obtained in this study highlight the effectiveness of near-infrared spectroscopy (NIR) combined with chemometric methods for the discrimination of tropical wood species. Based on a dataset of 864 samples from 12 different species, the model built on raw spectra achieved an overall classification accuracy of 97.80%, while applying the first derivative significantly improved this rate to 98.96%. This improvement confirms the usefulness of this preprocessing technique in correcting light scattering effects and enhancing the separation of closely related spectral signals, consistent with the observations of [2, 12], who demonstrated that the application of the Savitzky–Golay derivative improved the classification accuracy of several tropical species using an SVM model.

Several species, including Mahogany (Ac), Gmelina (Gm), Samba (Sa), Iroko (Ir), Teak (Te), and Bete (Be), were classified with 100% accuracy regardless of the preprocessing conditions, indicating a strong species-specific spectral signature. Others, such as Fraké (Fr), Dabema (Da), and Cedrela (Ce), showed notable improvement after the application of the first derivative, suggesting a reduction in spectral confusion with similar species.

Table 4
Classification of wood sawdust samples by cross-validation with models generated in PLS-DA, using a database both without and with preprocessing.

Species	Ac	Da	A	Sa	Gm	Ir	Te	Fr	Ce	He	Co	Be	Classified	% Classification	N Total
Untreated															
Ac	72												72	100.00	72
Da		69		2					1				69	95.83	72
A			69				1		1		1		69	95.83	72
Sa				72									72	100.00	72
Gm					72								72	100.00	72
Ir						72							72	100.00	72
Te							72						72	100.00	72
Fr				2				66	4				66	91.66	72
Ce		1			3				68				68	94.44	72
He					1					71			71	98.61	72
Co	2										70		70	97.22	72
Be												72	72	100.00	72
													845	97.80	864
1st Derivative															
Ac	72												72	100.00	72
Da		72											72	100.00	72
A			71				1						71	98.61	72
Sa				72									72	100.00	72
Gm					72								72	100.00	72
Ir						72							72	100.00	72
Te							72						72	100.00	72
Fr		1						71					71	98.61	72
Ce					2				70				70	97.22	72
He					2					70			70	97.22	72
Co	2		1								69		69	95.83	72
Be												72	72	100.00	72
													855	98.96	864

These findings are comparable to those of [13], who reported classification performance ranging from 95% to 99% depending on the surface quality (sawn, sanded, raw) during the identification of African woods using NIR spectroscopy. For instance, the work of [7] investigated the feasibility of using a portable near-infrared spectrometer combined with partial least squares analysis to identify five similar wood species of the *Cinnamomum* genus. Their study showed that models built from preprocessed spectra combining SNV and first derivative achieved identification accuracy above 95%. Similarly, the study by [14], which evaluated the use of Fourier Transform Near Infrared (FT-NIR) spectroscopy to distinguish fallen trees from the Lecythidaceae family, demonstrated the ability of FT-NIR spectroscopy to differentiate at the genus level within this family.

Compared to other approaches, particularly those using FTIR spectroscopy combined with LDA, KNN, or SVM algorithms, our results remain highly competitive. For example, [15] achieved correct classification rates ranging from 96% to 98% for Brazilian wood species under cross-validation, results very close to those observed in the present study.

More recently, studies integrating deep learning (LSTM, CNN) with NIR spectra have reached 97% accuracy for certain difficult-to-identify species, highlighting the emergence of new hybrid approaches combining spectroscopy and artificial intelligence [16].

In this study, the classification accuracy percentages of the PLS-DA models ranged from 0.90 to 0.99, which aligns with the results of numerous studies evaluating wood species classification [7, 16, 17]. The coefficients obtained during cross-validation for differentiating certain species varied between 0.86 and 0.94, similar to results reported by [2]. The improved performance of the PLS-R and PLS-DA models on wood sawdust can be explained by the reduction in particle size [18, 19].

The PLS-DA classification model highlighted the effectiveness of near-infrared (NIR) spectroscopy in identifying various types of sawdust originating from Ivorian carpentry workshops. Our findings underscore the potential of NIR spectroscopy, when combined with straightforward preprocessing techniques and supervised classification algorithms, for the rapid, non-destructive, and reliable identification of tropical wood species even in contexts characterized by high botanical diversity.

Table 5
Different class-wise metrics to assess classification performance from the confusion matrix obtained by cross-validation using untreated spectra.

Species	TP	FN	FP	Recall	Precision	Specificity	F1
Ac	72	0	2	1.000	0.973	0.997	0.986
Da	69	3	1	0.958	0.986	0.999	0.972
A	69	3	0	0.972	1.000	1.000	0.979
Sa	72	0	4	0.958	1.000	1.000	0.971
Gm	72	0	4	0.958	1.000	1.000	0.971
Ir	72	0	0	1.000	1.000	1.000	1.000
Te	72	0	1	0.986	1.000	1.000	0.993
Fr	66	6	0	0.872	0.943	0.995	0.930
Ce	68	4	4	0.944	1.000	1.000	0.971
He	71	1	0	0.986	1.000	1.000	0.993
Co	70	2	1	0.972	1.000	1.000	0.986
Be	72	0	0	1.000	1.000	1.000	1.000

Table 6
Different class-wise metrics to evaluate the classification performance from the confusion matrix obtained by independent validation using spectra processed with the first derivative.

Species	TP	FN	FP	Recall	Precision	Specificity	F1
Ac	72	0	2	0.972	0.972	0.972	1.000
Da	72	0	1	0.986	0.986	0.986	1.000
A	71	1	1	0.972	0.986	0.986	0.993
Sa	72	0	4	0.947	0.958	0.947	1.000
Gm	72	0	0	1.000	1.000	1.000	1.000
Ir	72	0	0	1.000	1.000	1.000	1.000
Te	72	0	0	1.000	1.000	1.000	1.000
Fr	71	1	0	0.986	1.000	1.000	0.993
Ce	70	2	0	0.972	1.000	1.000	0.986
He	70	2	0	0.972	1.000	1.000	0.986
Co	69	3	0	0.958	1.000	1.000	0.979
Be	72	0	0	1.000	1.000	1.000	1.000

This study offers a robust methodological foundation for wood traceability in both commercial and regulatory frameworks, while also opening the door to the integration of advanced machine learning approaches in more complex identification scenarios.

The Variable Importance in Projection (VIP) profile obtained from the PLS model using the untreated spectra in the 920–1650 nm spectral region reveals that the most influential wavelengths are mainly located between 1400 and 1500 nm, where VIP values exceed 2. These bands are associated with O–H and C–H overtone vibrations of aromatic and phenolic structures, which are characteristic of lignin and wood extractives. This indicates that variations in lignin and phenolic extractive contents are the dominant factors driving the spectral discrimination of the wood samples. In contrast, wavelengths between 1000 and 1200 nm, attributed to O–H and C–H vibrations of cellulose, exhibit moderate VIP values, suggesting a secondary contribution of polysaccharides to the overall variability. These results demonstrate that the near-infrared response of the studied Ivorian woods is primarily governed by lignin and extractives, which are known to play a major role in the natural durability of wood due to their hydrophobic and bioactive properties. Consequently, the NIR-based PLS model efficiently captures the chemical signatures related to the resistance of these woods to biological degradation [12].

The VIP profile obtained from the PLS model using first-derivative pre-treated spectra in the 920–1650 nm region highlights several highly influential spectral bands associated with aromatic and phenolic functional groups. The highest VIP values, observed around 1200, 1450, and 1600–1650 nm, correspond to C–H and O–H overtones of lignin and polyphenolic extractives [12]. These regions are characteristic of aromatic structures and phenolic hydroxyl groups commonly found in flavonoids, tannins, and quinonoid compounds.

The prominence of these bands indicates that the predictive performance of the NIR-PLS model is mainly driven by variations in lignin chemistry and secondary metabolites rather than by polysaccharides alone. This is particularly relevant for tropical hardwoods, which are known to be rich in bioactive extractives responsible for their natural durability and biological resistance. Consequently, the VIP distribution confirms that near-infrared spectroscopy, combined with chemometrics, provides a robust non-destructive approach for capturing the chemical signatures governing wood durability and bioactivity.

Table 7
Classification of wood sawdust samples by external validation using models generated in PLS-DA, with the database both without and with preprocessing.

Species	Ac	Da	A	Sa	Gm	Ir	Te	Fr	Ce	He	Co	Be	Classified	% Classification	N Total
Untreated															
Ac	18												18	100.00	18
Da		16		2									16	88.89	18
A			16				1				1		16	88.89	18
Sa				18									18	100.00	18
Gm					18								18	100.00	18
Ir						18							18	100.00	18
Te							18						18	100.00	18
Fr		1		1				15	1				15	83.33	18
Ce		2			2				14				14	77.78	18
He					1					17			17	94.44	18
Co	3										15		15	83.33	18
Be												18	18	100.00	18
Total													201	93.05	216
1st Derivative															
Ac	18												18	100.00	18
Da		17						1					17	94.44	18
A			16			2							16	88.89	18
Sa				18									18	100.00	18
Gm					18								18	100.00	18
Ir						18							18	100.00	18
Te							18						18	100.00	18
Fr		1						16	1				16	88.89	18
Ce					2			1	15				15	83.33	18
He					1					17			17	94.44	18
Co			1								17		17	94.44	18
Be												18	18	100.00	18
Total													206	95.37	216

Table 8
Different class-wise metrics to evaluate classification performance based on the confusion matrix obtained from independent validation using untreated spectra.

Species	TP	FN	FP	Recall	Precision	Specificity	F1
Ac	18	0	3	1.000	0.750	0.970	0.857
Da	16	2	3	0.889	0.941	0.995	0.914
A	16	2	0	0.889	0.889	0.990	0.889
Sa	18	0	3	1.000	1.000	1.000	1.000
Gm	18	0	3	1.000	1.000	1.000	1.000
Ir	18	0	0	1.000	1.000	1.000	1.000
Te	18	0	1	1.000	1.000	1.000	1.000
Fr	15	3	0	0.833	0.882	0.990	0.857
Ce	14	4	1	0.778	1.000	1.000	0.875
He	17	1	0	0.944	1.000	1.000	0.971
Co	15	3	1	0.833	1.000	1.000	0.909
Be	18	0	0	1.000	1.000	1.000	1.000

Table 9
Different class-wise metrics to evaluate the classification performance from the confusion matrix obtained by independent validation using spectra processed with the first derivative.

Species	TP	FN	FP	Recall	Precision	Specificity	F1
Ac	18	0	0	1.000	0.947	0.995	0.973
Da	17	1	1	0.944	1.000	1.000	0.971
A	16	2	1	0.889	1.000	1.000	0.941
Sa	18	0	0	1.000	1.000	1.000	1.000
Gm	18	0	3	1.000	1.000	1.000	1.000
Ir	18	0	0	1.000	1.000	1.000	1.000
Te	18	0	2	1.000	1.000	1.000	1.000
Fr	16	2	2	0.889	0.941	0.995	0.914
Ce	15	3	1	0.833	0.882	0.990	0.857
He	17	1	0	0.944	1.000	1.000	0.971
Co	17	1	0	0.944	1.000	1.000	0.971
Be	18	0	0	1.000	1.000	1.000	1.000

5. Conclusion

Near-infrared spectroscopy, a non-destructive technique, was used to identify wood sawdust from carpentry workshops in Yamoussoukro (Côte d'Ivoire). The use of chemometric tools facilitated the interpretation of the spectra and allowed the identification of patterns among the studied wood sawdust species. PLS-DA models enabled the differentiation of sawdust from 12 Ivorian wood species. Some species showed high classification accuracy according to the developed models, while others showed less effective results. The PLS-DA model, combined with NIR spectroscopy, can be considered a highly promising tool for valorizing wood sawdust, which is often perceived as waste in Côte d'Ivoire. The results of this study are specific to Côte d'Ivoire, and the conclusions would remain equally reliable with a larger number of samples, provided that appropriate calibration is used. Near-infrared spectroscopy, combined with chemometric tools which proved effective in this study, holds great promise for future research in the wood sector, particularly for distinguishing between Ivorian wood species.

Data availability

The NIR data used in this study are available on https://zenodo.org/ with the link https://doi.org/10.5281/zenodo.14996051.

Acknowledgments

We thank the Société de Développement des Forêts (SODEFOR) and INPROBOIS for supplying plant material. We would also like to thank the Doctoral School of INP-HB.

References

[1] W. Kouamé Kra, Côte d'Ivoire : les aires protégées entre politique de conservation contrastée et réinterprétation sociale, Études caribéennes (2019). https://doi.org/10.4000/etudescaribeennes.17124

[2] J.L.L. N'guessan, B.F. Niamké, H.A.A. Ouattara *et al.*, *Identification of wood species from Côte d'ivoire forest using near-infrared spectroscopy*, Wood Sci. Technol. 59 (2025) 72. https://doi.org/10.1007/s00226-025-01677-z

[3] A. Watanabe, S. Morita, S. Kokot *et al.*, *Drying process of microcrystalline cellulose studied by at-*

tenuated total reflection IR spectroscopy with two-dimensional correlation spectroscopy and principal component analysis, J. Mol. Struct. 799 (2006) 102–110.
https://doi.org/10.1016/j.molstruc.2006.03.018

[4] D.C. Silva, T.C.M. Pastore, L.F. Soares *et al.*, *Determination of the country of origin of true mahogany (Swietenia macrophylla King) wood in five Latin American countries using handheld NIR devices and multivariate data analysis*, Holzforschung 72 (2018) 521–530.
https://doi.org/10.1515/hf-2017-0160

[5] P.R.G. Hein, *Estimating Shrinkage, Microfibril Angle and Density of Eucalyptus Wood Using near Infrared Spectroscopy*, J. Near Infrared Spectrosc. 20 (2012) 427–436.
https://doi.org/10.1255/jnirs.1005

[6] R.G. Brereton, G.R. Lloyd, *Partial least squares discriminant analysis: taking the magic away*, J. Chemom. 28 (2014) 213–225.
https://doi.org/10.1002/cem.2609

[7] X. Pan, J. Qiu, Z. Yang, *Identification of Five Similar Cinnamomum Wood Species Using Portable Near-Infrared Spectroscopy*, Spectroscopy 37(6) (2022) 16–23, 49.
https://doi.org/10.56530/spectroscopy.zg7089n4

[8] X. Xue, Z. Chen, H. Wu *et al.*, *Identification of Guiboutia species by NIR-HSI spectroscopy*, Sci. Rep. 12 (2022) 11507.
https://doi.org/10.1038/s41598-022-15719-0

[9] X. Pan, Z. Yu, Z. Yang, *A Multi-Scale Convolutional Neural Network Combined with a Portable Near-Infrared Spectrometer for the Rapid, Non-Destructive Identification of Wood Species*, Forests 15 (2024) 556.
https://doi.org/10.3390/f15030556

[10] X. Xue, Z. Chen, H. Wu *et al.*, *Identification of Eight Pterocarpus Species and Two Dalbergia Species Using Visible/Near-Infrared (Vis/NIR) Hyperspectral Imaging (HSI)*, Forests 14 (2023) 1259.
https://doi.org/10.3390/f14061259

[11] B. Leblon, O. Adedipe, G. Hans *et al.*, *A review of near-infrared spectroscopy for monitoring moisture content and density of solid wood*, For. Chron. 89 (2013) 595–606.
https://doi.org/10.5558/tfc2013-111

[12] E.N. Mezui, L. Brancheriau, S. Ikogou *et al.*, *Linking near-infrared spectroscopy of 25 tropical Gabonese hardwoods to tree ecological temperament*, Bois For. Trop. 362 (2025) 1–14.
https://doi.org/10.19182/bft2025.362.a37589

[13] S. Huancas, D.T. Medeiros, T.L. Dias *et al.*, *Impact of surface quality on the identification of tropical wood species using benchtop and portable NIR instruments*, Research Square (2024) 1–19.
https://doi.org/10.21203/rs.3.rs-4681106/v1

[14] C. Eugenio Da Silva, C.S. Nascimento, J.A. Freitas *et al.*, *Alternative identification of wood from natural fallen trees of the Lecythidaceae family in the Central Amazonian using FT-NIR spectroscopy*, Int. Forest. Rev. 26 (2024) 29–44.
https://doi.org/10.1505/146554824838457844

[15] E. Jesus, T. França, C. Calvani *et al.*, *Making wood inspection easier: FTIR spectroscopy and machine learning for Brazilian native commercial wood species identification*, RSC Adv. 14 (2024) 7283–7289.
https://doi.org/10.1039/d4ra00174e

[16] Z. Wan, H. Yang, J. Xu *et al.*, *BACNN: Multi-scale feature fusion-based bilinear attention convolutional neural network for wood NIR classification*, J. For. Res. 35(4) (2024) 1–13.
https://doi.org/10.1007/s11676-023-01652-z

[17] L.F. Soares, D.C. Silva, M.C.J. Bergo *et al.*, *Avaliação de espectrômetro NIR portátil e PLS-DA para a discriminação de seis espécies similares de madeira amazônicas*, Quim. Nova 40(4) (2017) 418–426.
https://doi.org/10.21577/0100-4042.20170014

[18] A.C. Raobelina, G. Chaix, A.R. Razafimahatratra *et al.*, *Use of a Portable Near Infrared Spectrometer for Wood Identification of Four Dalbergia Species from Madagascar*, WFS 55 (2023) 4–17.
https://doi.org/10.22382/wfs-2023-03

[19] L. Servant, G. Le Bourdon, T. Buffeteau, Comprendre la spectroscopie infrarouge : principes et mise en oeuvre, Photoniques 53 (2011) 68–73.
https://doi.org/10.1051/photon/20115368