

Classification of four wood species using terahertz time-domain spectroscopy THZ-TDS

Jimmy Frank Oblitas-Cruz¹ <https://orcid.org/0000-0001-7652-6672>*

Glicerio Eduardo Torres-Carranza² <https://orcid.org/0000-0001-5210-0554>

¹Universidad Privada del Norte. Facultad de Ingeniería. Centro de Investigación Avanzada en Agroingeniería. Cajamarca, Perú.

²Universidad Nacional de Cajamarca. Facultad de Ciencias Agrarias. Departamento de agronomía. Cajamarca, Perú.

*Corresponding author: Jimmy.oblitas@upn.edu.pe

Abstract:

Terahertz time-domain spectroscopy (THz-TDS) combined with chemometric techniques was investigated as a non-destructive method for species classification in wood samples from four different species. For this purpose, samples were introduced into the THz-TDS spectrometer chamber purged with dry nitrogen, and spectra were collected at wave intervals from 0,1 THz to 10 THz to evaluate the samples. It was obtained that in the range between 0,1 THz and 1,7 THz, there is higher differentiation of the samples, being this the best range to generate classification models. After the chemometric analysis, the Support Vector Machine algorithm achieved an accuracy rate of 91 % in classifying the four wood species. Finally, it was demonstrated that THz spectroscopy can be used for quantitatively complex natural organic materials such as wood.

Keywords: Chemometrics analysis, multivariate analysis, non-destructive wood analysis, terahertz time-domain spectroscopy, wood species identification.

Received: 24.03.2024

Accepted: 15.03.2026

Introduction

Wood is a natural material of economic importance for countries with large forest areas. Different wood species exhibit significant variations in properties, prices, and applications, basically due to the great diversity of wood species (Wang *et al.* 2023a), which generates fluctuations in their physical properties, such as moisture content, grain

direction, and density. Determining these values for each species involves assays that are generally time-consuming and rely primarily on macroscopic observation that require highly trained inspectors, and have low detection efficiency, or processes such as microscopic cutting or DNA barcoding (Yu *et al.* 2016) that are costly and damage samples.

Accurate wood species identification is essential for sustainable forest management and the control of illegal timber trade. However, conventional methods based on anatomical or chemical analyses are often time-consuming, require specialized expertise, and may involve destructive sampling. Therefore, there is an increasing demand for rapid and non-destructive analytical approaches. Spectroscopy can classify wood by measuring light absorption based on different chemical compounds in the wood. The absorption spectrum of wood is unique to each tree species, so by comparing the absorption spectrum of a wood sample with a database of known wood species, it is possible to identify the wood species (Bensalem *et al.* 2018).

In this context, it is necessary to explore new methods for carrying out a non-destructive and fast classification of wood. The literature shows the use of spectral analysis for this purpose, such as near-infrared spectroscopy (NIR) (Medeiros *et al.* 2024), hyperspectral imaging (Chen and Li 2020), and RAMAN spectroscopy (Kusnierek *et al.* 2024), among others. These spectral methods have been evolving based on the need for a non-destructive evaluation of wood to promote its widespread use in the industry. (Tanaka *et al.* 2014).

Currently, there is a study gap in the spectral evaluation range, which is in the terahertz length (Arteaga *et al.* 2021), that has generated interest in exploring the potential of this electromagnetic range in wood classification. According to (Trafela *et al.* 2013), wood

components such as cellulose, hemicellulose, and lignin have vibrational energy levels that can be detected in the THz range.

THz spectroscopy has been widely used in various fields linked to biological tissues (Cruz 2020), including the detection of some forest species (Bensalem *et al.* 2018), where superiority to classify timber species is reported. THz time-domain spectroscopy allows the analysis of wood dielectric properties (Wang *et al.* 2023b), being able to improve the accuracy of wood identification studies.

The objective of this study was to evaluate the potential of terahertz time-domain spectroscopy (THz-TDS) combined with chemometric analysis for the spectral discrimination and classification of four wood species. Specifically, the research aimed to assess whether the spectral signatures obtained in the THz region can effectively differentiate wood species of technological and commercial relevance, thereby contributing to the development of rapid and non-destructive approaches for wood identification. In this context, the study also explores the methodological advantages of THz spectroscopy, highlighting its potential as an emerging analytical tool for non-destructive wood testing and species classification.

Materials and methods

Samples

Wood samples from four wood species from the North of Peru were selected: copaiba (*Copaifera officinalis* L.), catahua (*Hura crepitans* L.), cedar (*Cedrela odorata* L.), and

radiata pine (*Pinus radiata* D.Don). The samples were selected and taken to a laboratory where cubic pieces 10 cm long, 10 cm wide, and 10 cm thick were extracted and then divided into 10 samples of 1 cm thickness each.

Physical Characterization of Wood Samples

The determination of the physical properties of the wood samples was carried out following the procedures established in the Peruvian technical standards. Basic density was determined according to NTP 251.011:2014, which specifies the method for calculating the ratio between the oven-dry mass of the specimen and its saturated volume. The shrinkage properties were evaluated in accordance with NTP 251.012:2015, through which tangential, radial, and volumetric shrinkage were determined based on the dimensional variations recorded between the saturated and oven-dry states of the specimens. The specimens were handled and conditioned according to the requirements of the corresponding standards to minimize experimental variability and to guarantee the reliability of the physical property determinations.

When significant differences were detected ($p < 0.05$), Tukey's multiple comparison test was applied to identify specific differences between the mean values of the species. This post-hoc test allowed the classification of the wood species into statistically homogeneous groups according to their physical properties.

Data acquisition with THZ-TDS spectrometer

The cubic samples were introduced into the THZ-TDS spectrometer chamber. The latter was purged with dry nitrogen to avoid air humidity disturbances during the measurement. A TeraView TeraPulse 4000 stationary terahertz spectrometer was used. Spectra were collected in time intervals from 1 μs to 25 μs with a time resolution of 5 ps; a terahertz light source with a wavelength of 0,1 THz to 10 THz was used to evaluate the samples. Subsequently, the recorded time-domain signals were transformed into the frequency domain through a Fast Fourier Transform (FFT) procedure. The extraction and processing of THz spectral data followed the methodology described by (Cruz 2020), where time-delay signals are converted into frequency-dependent spectral information for further analysis.

Statistical analysis

The data were analyzed using Matlab 2023a software (The MathWorks Inc. 2023), where signal processing algorithms, machine learning algorithms, statistical analysis techniques, and spectra visualizations were used. The samples taken were in the time domain and were processed using a Fast Fourier Transform to obtain spectra in the frequency and absorbance domain. A principal component analysis with 24 machine learning algorithms were used using the Classification Learner package from matlab. The results were compared to a reference data set to determine the classification accuracy; the reference data are the spectra taken when the chamber is empty. The summary of the methodology is shown in Figure 1.

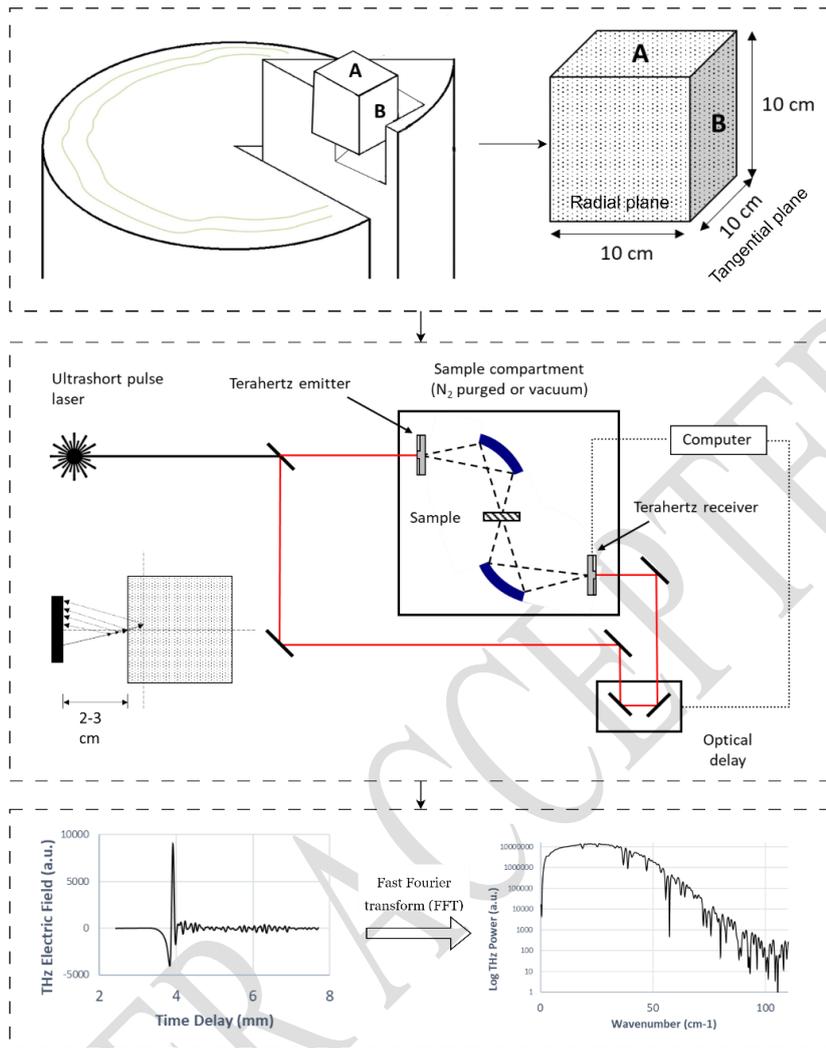


Figure 1: Process of obtaining the spectral image of the samples and transformation of the signal from time-domain to frequency-domain.

Results and discussion

Sample characterization

Figure 2 shows the physical properties of the wood samples studied. In addition, a Tukey's multiple comparison test was performed to observe whether there are significant changes between species.

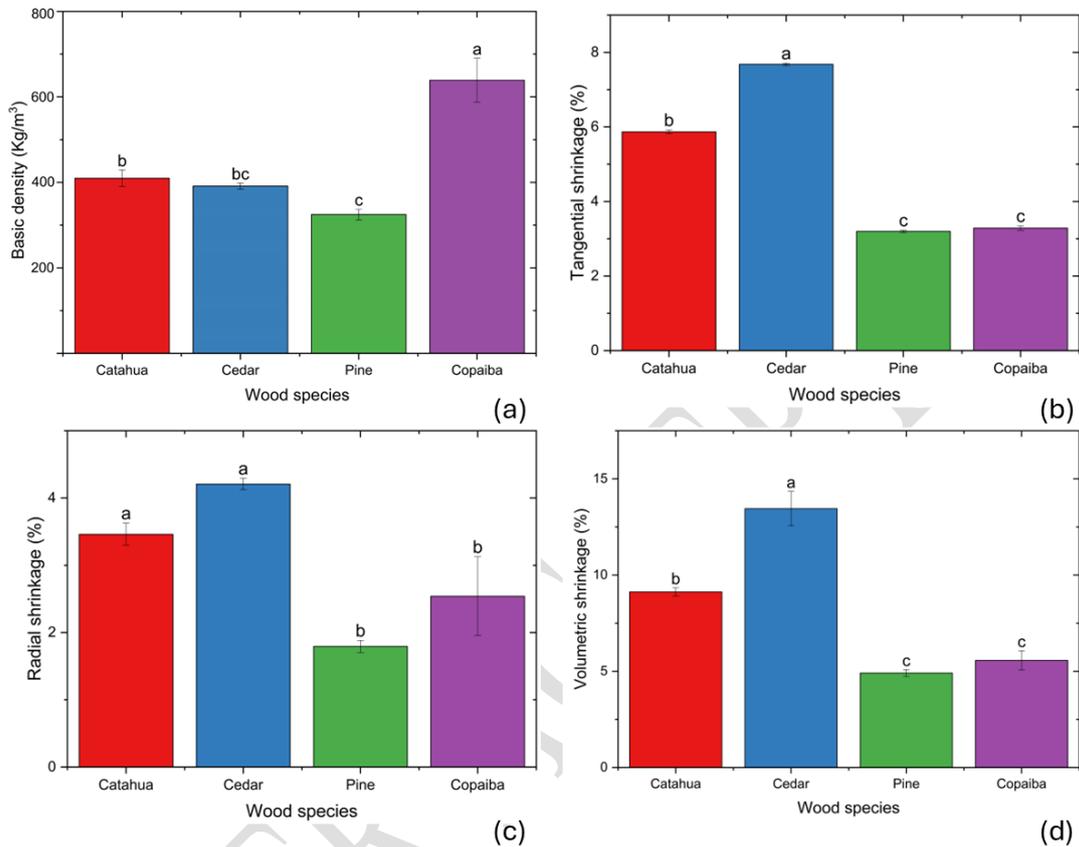


Figure 2: Physical properties (a) Basic density, (b) Tangential shrinkage, (c) Radial shrinkage, and (d) Volumetric shrinkage.

The samples used were obtained under controlled conditions, using a drying oven at 22 °C and a relative humidity of 66 %, with the aim of reaching a moisture content of 12 % (Moya *et al.* 2025). This property is very variable, helping to classify wood as very light, light, semi-heavy, heavy, and very heavy. According to these scales, cedar, catahua, and pine would be light and copaiba semi-heavy. Wood is generally a porous material. The amount of solid substance in a volume of wood determines its strength properties, such as workability and drying. The mass and volume of a piece of wood vary significantly

according to the moisture content and depend on the condition under which it is obtained. The saturation point of the fiber and anhydrous mass determine the basic density. Under these characteristics, radiata pine (*Pinus radiata* D.Don) wood has an average basic density of 390 kg/m³, catahua (*Hura crepitans* L.) 410 kg/m³, cedar (*Cedrela odorata* L.) 420 kg/m³, and copaiba (*Copaifera officinalis* L.) 610 kg/m³.

Another physical property is the shrinkage that wood undergoes as it loses water below the saturation point of the fibers. These are radial, tangential, and volumetric. It is produced due to the anisotropy characterizing wood, which varies its dimensions and, consequently, its volume. Hence, its evaluation is specified in three types of shrinkage (Monteoliva 2010). Specifically, broadleaved wood is heterogeneous. The wood of the three broadleaved species in this study presents diffuse porosity, solitary and multiple vessels, and pores; the parenchyma is diffuse in aggregates (catahua), diffuse, vasicentric, and marginal (cedar and copaiba), and with secretory canals and vasicentric paratracheal (copaiba); unstratified rays in the tangential section (catahua and copaiba), fine rays in the tangential face, and unstratified in the radial face (cedar); in its inclusions, it presents tyloses in the pores (catahua), dark brown resins (cedar) and abundant inclusions of resin in the vessels and the limit of the growth rings.

Classification using THZ-TDS spectroscopy

Figure 3 shows a graphical representation of the analysis performed for the THZ-TDS values. On the left, with the index "a," the spectral image of each tree species sample

copaiba (*Copaifera officinalis* L.), radiata pine (*Pinus radiata* D.Don), catahua (*Hura crepitans* L.), cedar (*Cedrela odorata* L.) is shown, and, on the right, the extraction of the spectra in the time-domain with the index "b." These values were treated with a Fast Fourier Transform (FFT) to achieve a graphical representation of the terahertz spectroscopy data obtained in the time-domain to classify wood species of pine, copaiba, cedar, catahua shown with the index "c." These values are then transformed into indicators that allow us a classification based on their absorbance (Figure 4).

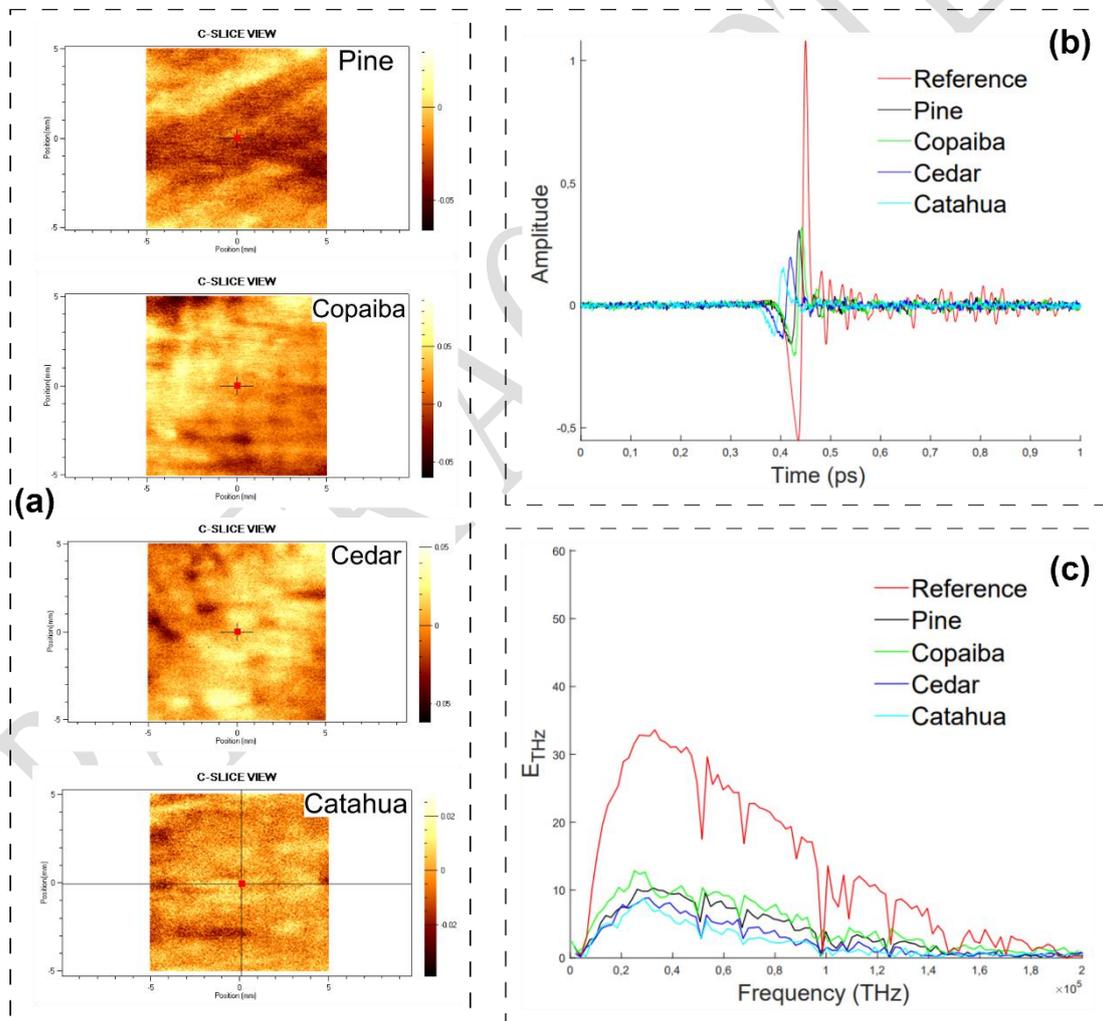


Figure 3: (a) Time-domain images, (b) Measured THz pulses from the reference and the sample, (c) Corresponding amplitude spectra of the measured reference pulse and the sample pulses.

Figure 4 shows the absorbance calculated for the four different wood species. It was evaluated from 0,1 to 8 THz, showing the main differences in the intensity and location of the absorption peak, which are more noticeable in the frequency range of 0,1 to 1,7 THz. This range, where the difference between the samples is more notorious, coincides with the work done by (Yun *et al.* 2024), who found a similar range when evaluating hardwoods and softwoods. The absorption coefficient spectra of the four wood species showed more overlap beyond 1,8 THz, and the differences were insignificant. While the four wood species show similarities in the peaks and trend of their absorbance curves in the 0,1 to 1,7 THz range, the intensity varied significantly.

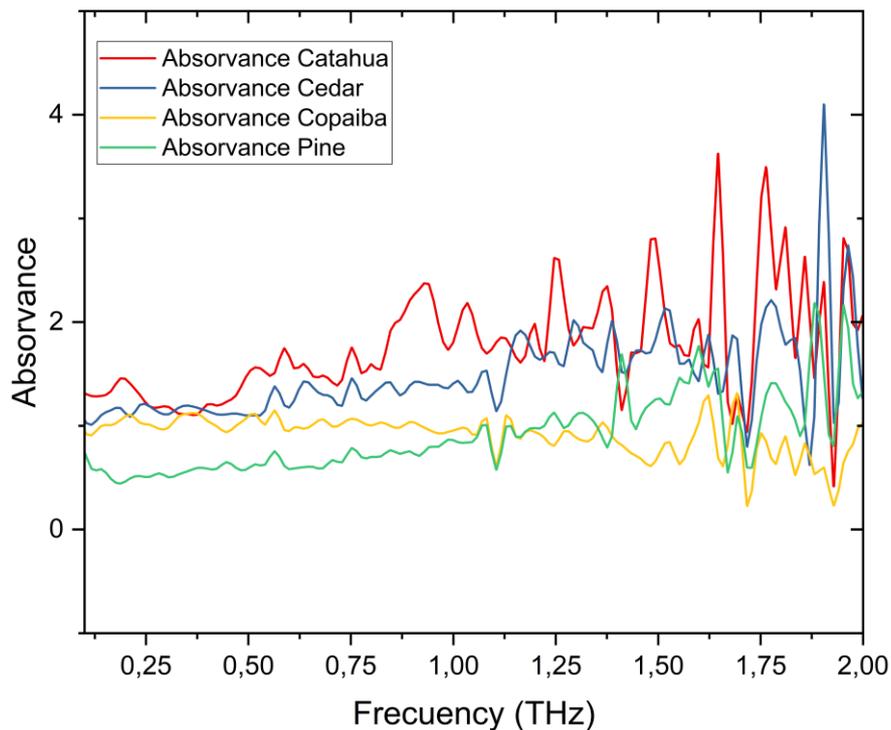


Figure 4: Absorbance spectrum in the THz range.

This differentiation between the spectra in the range 0,1 THz to 1,7 THz of the four wood species may be due to several reasons. One of them may be related to the anatomical

characteristics of the wood in each wood species, especially with the homogeneity of the wood. For example, pine has a very homogeneous curve since approximately 80 % of the cells are of a single type, the tracheids; copaiba, on the other hand, has a similar behavior, although to a lesser degree due to its density, which makes its xylem cells smaller and with thick walls, while cedar has a more heterogeneous xylem, with various types of cells such as vessels, fibers, and parenchyma. Catahua presents a particular case. Its density is very similar to cedar as its xylem anatomy. However, its behavior is different. This may be due to the high presence of latex-type extractives, or the possibility of making new measurements should be explored to rule out possible errors caused by the composition of this type of wood.

This type of appreciation has been reported in research using THz spectroscopy to see differences between woods, relating them to a difference in composition between alkaline and hydrolytic lignin (Trafela *et al.* 2013). Likewise, it is observed that there is a direct relationship between wood density and the THz readings found. In general, the lower the density, the earlier absorbance readings start in the coordinate graph. Therefore, the highest absorbance calculated is generated by catahua followed by cedar, which has a very similar average density, and pine is the one with the lowest absorbance, possibly due to the lower basic density. Other studies on wood density and its relationship with THz have been reported and determined that wood scatters the terahertz wave with a wavelength close to the dimensions of the cross-section of the pores of the wood (Tanaka *et al.* 2014). Likewise, (Inagaki *et al.* 2014) showed that, in the range from 0,1 THz to 0,2 THz, wood samples had a strong correlation with their density and moisture content. The 0,1 - 1,7 THz spectrum was selected for further data processing and modeling. Preprocessing of the absorbance data was performed using the Savitsky-Golay method,

and a principal component analysis was performed, as shown in Figure 5, where with only 2 PCs the total explanation of the data variance is obtained.

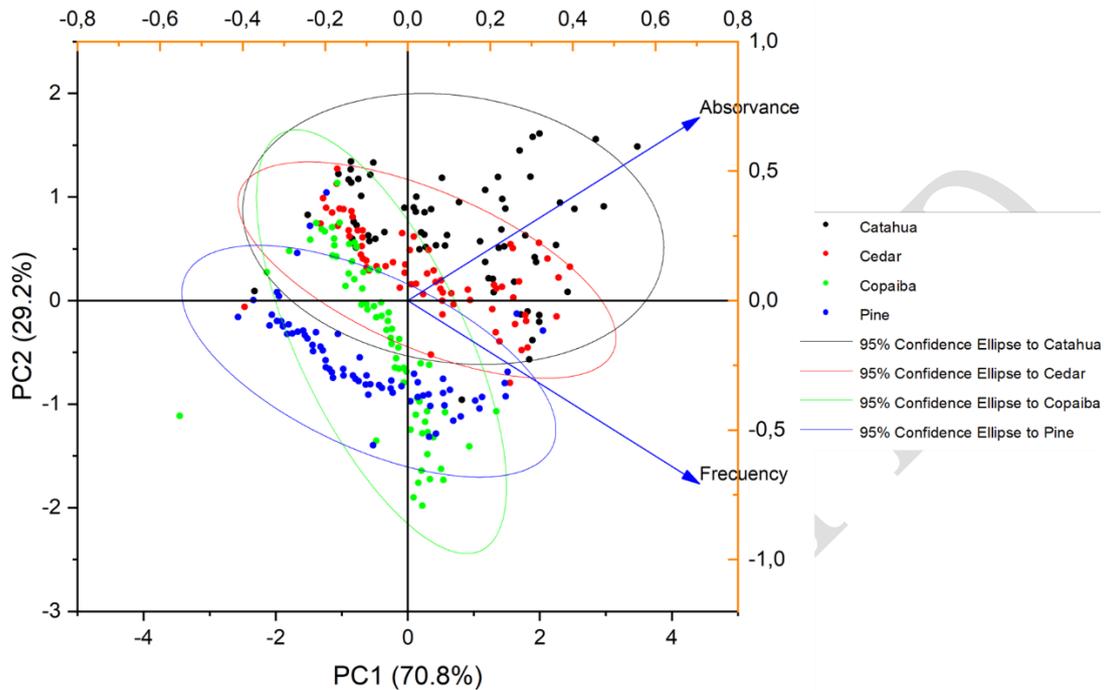


Figure 5: Principal component analysis (PCA)

Principal Component Analysis (PCA) revealed clear patterns in the spectral behavior of the studied wood species (Figure 5). The first two principal components explained 100% of the total variance, with PC1 accounting for 70.8% and PC2 explaining 29.2%, indicating that most of the variability in the dataset can be described in a two-dimensional space. The score plot shows a partial separation among the four wood species. PC1 is mainly associated with frequency, distinguishing Cedar and Catahua on the positive side from Pine and Copaiba, which tend to cluster toward negative values. PC2 is primarily related to absorbance, with Catahua samples located mostly in the positive region and Copaiba in the negative region.

The 95% confidence ellipses illustrate the dispersion and grouping tendencies of each species. Although partial overlap among species is observed, particularly between Cedar and Catahua, the clusters show distinguishable tendencies. Pine forms a relatively compact cluster in the lower region of the plot, while Copaiba exhibits a broader dispersion, reflecting higher intra-species variability. Overall, the PCA results demonstrate that the spectral variables frequency and absorbance contribute significantly to the discrimination of the studied wood species.

To determine the classification level, 24 models were tested and generated using Matlab 2023a. Table 1 shows the 5 models with the highest accuracy, achieving values from 85 % to 91 % accuracy. The performance of the classification models was evaluated using a supervised machine learning framework implemented in MATLAB using the Classification Learner App. To ensure a robust comparison among models, a k-fold cross-validation procedure ($k = 5$) was applied to the dataset. In this approach, the dataset was randomly partitioned into five equal subsets; during each iteration, four subsets were used for model training while the remaining subset was used for validation. This process was repeated five times so that each subset served once as the validation set. The final classification accuracy reported in Table 1 corresponds to the average accuracy obtained across the five folds, which reduces the risk of overfitting and provides a more reliable estimate of model generalization performance.

Table 1: Classification models.

| Model | Hyperparameters | Accuracy (%) |
|--------------------------|--|--------------|
| Support vector machine | Kernel function: Cubic | 91 |
| Support vector machine | Kernel function: Quadratic | 90 |
| Narrow neural network | First Layer size:10 Activation: ReLU | 90 |
| Bilayered neural network | Number of fully connected layers: 2 First Layer size:10 Second Layer size:10 Activation: ReLU | 88 |
| K-nearest neighbor | Distance metric: Euclidian Distance weight: Squared inverse | 85 |

The THz-TDS approach successfully identified the four different wood species with the highest accuracy rate of 91 %. The Support Vector Machine (SVM) algorithm yielded the best results. Similar reports, such as (Yun *et al.* 2024) using THz on softwood samples, found SVM to be the best algorithm for classification.

According to the results obtained using THz technology, the high potential to achieve effective recognition of wood species can be seen because wood is a composite material, and the structure of its wood fibers has a strong influence on its properties. Studying the structure of wood fibers is crucial for comprehending the nature of wood and its effects on the wood industry. A thorough understanding of the structure of wood fibers can help identify areas for improvement in their characteristics, such as performance, strength, and stiffness, among other key factors. In general, wood classification is an essential tool that can be used to improve the quality, sustainability, and usability of wood, so results such as those obtained in the research become crucial inputs for future applications.

Conclusions

Terahertz time-domain spectroscopy was used to measure the spectrum in the range between 0,1 THz and 8 THz of wood samples of 4 wood species. It was found that the range between 0,1 THz and 1,7 THz is where there is a higher differentiation of the samples, being the best range to generate adequate classification models for these four wood species since they contain useful quantitative information, especially in samples such as wood with low water content, which are often considered transparent for THz. Likewise, after the chemometric analysis, it can be observed that classification models can be generated, reaching 91 % accuracy using the SVM algorithm. Finally, it can be affirmed that THz spectroscopy can be used for the quantitative characterization of complex natural organic materials, even when it is still possible to improve the levels of accuracy by using more classification models and including prediction models that relate them to the characteristics of each type of wood.

This study demonstrates the capability of Terahertz Time-Domain Spectroscopy as a non-destructive analytical technique for the characterization and discrimination of wood species. Spectral measurements acquired in the 0,1–8 THz range revealed that the frequency interval between 0,1 and 1,7 THz contains the most relevant spectral information for differentiating the analyzed samples. This region exhibited the highest discriminatory potential, highlighting its suitability for developing robust classification models, particularly for lignocellulosic materials such as wood with low moisture content, which are often considered relatively transparent to THz radiation.

The integration of THz spectroscopy with chemometric analysis enabled the development of reliable classification models. Among the evaluated algorithms, the model based on Support Vector Machine achieved the best performance, reaching an

accuracy of approximately 91 % in the classification of the four studied wood species. These results confirm that THz spectral signatures contain sufficient structural and compositional information to enable the discrimination of complex natural materials.

From a broader perspective, the findings highlight the potential of THz spectroscopy combined with machine learning approaches as a rapid and environmentally friendly alternative to conventional wood identification techniques. This approach could support applications in forestry management, wood industry quality control, and timber traceability. Such developments could further enhance the accuracy, robustness, and practical applicability of THz-based analytical methods for lignocellulosic materials.

Authorship contributions

J. O.: Conceptualization, formal analysis, funding acquisition, methodology, writing - review & editing, supervision, funding acquisition, project administration. G. T.: Methodology, investigation.

Conflicts of interest

The authors declare there are no conflicts of interest for each author.

References:

- Arteaga, H.; León-Roque, N.; Oblitas, J. 2021.** The frequency range in THz spectroscopy and its relationship to the water content in food: A first approach. *Scientia Agropecuaria* 12(4): 625-634. <https://doi.org/10.17268/sci.agropecu.2021.066>
- Bensalem, M.; Sommier, A.; Mindeguia, J.C.; Batsale, J.C.; Pradere, C. 2018.** Terahertz Measurement of the Water Content Distribution in Wood Materials. *Journal of Infrared, Millimeter, and Terahertz Waves* 39(2): 195-209. <https://doi.org/10.1007/s10762-017-0441-7>

- Chen, J.; Li, G. 2020.** Prediction of moisture content of wood using Modified Random Frog and Vis-NIR hyperspectral imaging. *Infrared Physics & Technology* 105. e103225. <https://doi.org/10.1016/j.infrared.2020.103225>
- Cruz, J.O. 2020.** Terahertz Time-domain Spectroscopy (THz-TDS) for classification of blueberries according to their maturity. In: *Proceedings of the 2020 IEEE Engineering International Research Conference (EIRCON 2020)*. IEEE: Lima, Peru. 26-30 October 2020. <https://doi.org/10.1109/EIRCON51178.2020.9254046>
- Instituto Nacional de Calidad (INACAL). 2014. NTP 251.011:2014.** Madera. Determinación de la densidad básica. Lima, Perú: Instituto Nacional de Calidad.
- Instituto Nacional de Calidad (INACAL). 2015. NTP 251.012:2015.** Madera. Determinación de la contracción radial, tangencial y volumétrica. Lima, Perú: Instituto Nacional de Calidad.
- Inagaki, T.; Ahmed, B.; Hartley, I.D.; Tsuchikawa, S.; Reid, M. 2014.** Simultaneous prediction of density and moisture content of wood by terahertz time domain spectroscopy. *Journal of Infrared, Millimeter, and Terahertz Waves* 35(11): 949-961. <https://doi.org/10.1007/s10762-014-0095-7>
- Kusnierek, K.; Woznicki, T.; Treu, A. 2024.** Quality control of wood treated with citric acid and sorbitol using a handheld Raman spectrometer. *Journal of Cleaner Production* 434. e139925. <https://doi.org/10.1016/j.jclepro.2023.139925>
- Medeiros, D.T. de; Gomes, J.N.N.; Batista, F.G.; Mascarenhas, A.R.P.; Pimenta, E.M.; Chaix, G.; Hein, P.R.G. 2024.** Estimation of the basic density of *Eucalyptus grandis* wood chips at different moisture levels using benchtop and handheld NIR instruments. *Industrial Crops and Products* 209. e117921. <https://doi.org/10.1016/j.indcrop.2023.117921>
- Monteoliva, S.E. 2010.** La madera como material: Estructura anatómica y propiedades. Editorial Universidad Nacional de La Plata: La Plata, Argentina. <https://ri.conicet.gov.ar/handle/11336/136615>
- Moya, R.; Tenorio, C.; Villalobos-Barquero, V.; Meza-Montoya, A. 2025.** Variación de las propiedades físicas de la madera y efecto de las variables dasométricas en árboles de *Ochroma pyramidale* que crecen en plantaciones. *Heliyon* 11(1). e41210. <https://doi.org/10.1016/j.heliyon.2024.e41210>
- Tanaka, S.; Shiraga, K.; Ogawa, Y.; Fujii, Y.; Okumura, S. 2014.** Applicability of effective medium theory to wood density measurements using terahertz time-domain spectroscopy. *Journal of Wood Science* 60(2): 111-116. <https://doi.org/10.1007/s10086-013-1386-7>
- The MathWorks Inc. 2023.** MATLAB: numerical computing environment and programming language. The MathWorks Inc.: Natick, Massachusetts, USA. <https://www.mathworks.com>
- Trafela, T.; Mizuno, M.; Fukunaga, K.; Strlič, M. 2013.** Quantitative characterisation of historic paper using THz spectroscopy and multivariate data analysis. *Applied Physics A* 111(1): 83-90. <https://doi.org/10.1007/s00339-012-7525-y>
- Wang, Y.; Gao, R.; Ma, L.; Kang, K.; Wang, C.; Guo, Y.; Ge, X. 2023a.** Analysis of the application status of terahertz technology in forestry. *European Journal of Wood and Wood Products* 82: 561-578. <https://doi.org/10.1007/s00107-023-02025-3>
- Wang, Y.; He, Y.; Wang, Z.; Avramidis, S. 2023b.** Information fusion technology for terahertz spectra and hyperspectral imaging in wood species identification. *European Journal of Wood and Wood Products* 82: 579-589. <https://doi.org/10.1007/s00107-023-02027-1>

Yu, M.; Liu, K.; Zhou, L.; Zhao, L.; Liu, S. 2016. Testing three proposed DNA barcodes for the wood identification of *Dalbergia odorifera* T. Chen and *Dalbergia tonkinensis* Prain. *Holzforschung* 70(2): 127-136. <https://doi.org/10.1515/hf-2014-0234>

Yun, X.D.; Wang, Y.; Ma, W.J.; Zhao, L. 2024. Wood Recognition Based on Terahertz Spectrum and Hyperspectral Technology. *Journal of Applied Spectroscopy* 90(6): 1422-1428. <https://doi.org/10.1007/s10812-024-01680-5>

PAPER ACCEPTED