

Discrimination of caatinga species based on bark using near infrared spectroscopy

Silvana Nisgoski^{1*}, Helena Cristina Vieira², Joielan Xipaia dos Santos¹,
Stephanie Hellen Barbosa Gomes¹, Rosimeire Cavalcante dos Santos³,
Graciela Inés Bolzon de Muñiz¹

¹Federal University of Paraná, Department of Forest Engineering and Technology, Curitiba, PR, Brazil

²Federal Rural University of Pernambuco, Department of Forest Science, Recife, PE, Brazil

³Federal University of Rio Grande do Norte, Jundiá School of Agriculture, Macaíba, RN, Brazil

TECHNOLOGY OF FOREST PRODUCTS

ABSTRACT

Background: The Caatinga biome has high deforestation rate, so the correct identification of species is important to conserve resources. The objective of this study was to evaluate the potential of NIR spectroscopy to distinguish tree barks from eight species from the Caatinga biome based on the development of multivariate models. Three trees of each species were felled, and the trunk was cut at six positions to obtain bark sample discs: 0%, diameter at breast height (DBH) (1.30 m from ground), 25%, 50%, 75% and 100% of commercial height. Spectra were collected with resolution of 4 cm⁻¹ and wavenumber ranging from 10 000 to 4 000 cm⁻¹ using a probe with 2 mm aperture. All discs obtained from the six positions were approximately 5 mm from the probe, and 24 spectra were collected from each disc, for a total of 144 per tree and 432 per species. Classification methods were based on all spectra and only the DBH position, by applying linear discriminant analysis, support vector machine and k-nearest neighbors (K-NN).

Results: Better results were obtained with K-NN and first derivative spectra, with accuracy of 0.91 (all tree positions) and 0.85 (only DBH). NIR spectroscopy with multivariate analysis has potential to discriminate Caatinga species based on spectra of bark samples.

Conclusion: The use of near infrared in forest can confirm the correct species before cut on forest management, contributing to conservation of Caatinga resources and an adequate use of species with high aggregated value.

Key words: Forest inspection; multivariate analysis; native woods; NIR.

HIGHLIGHTS

It is possible to discriminate Caatinga species based on bark spectra.
There is influence of tree height on the near-infrared spectra obtained from the bark.
Better results in discriminating the bark of Caatinga species were observed using K-NN.
In K-NN, the processed data into first derivative presented higher precision values.

NISGOSKI, S.; VIEIRA, H. C.; SANTOS, J. X.; GOMES, S. H. B.; SANTOS, R. C.; MUNIZ, G. I. B. Discrimination of Caatinga species based on bark using near infrared spectroscopy. CERNE, v.31, e-103490, 2025. doi: 10.1590/01047760202531013490

*Corresponding author: silvana.ufpr@gmail.com
Scientific Editor: Paulo Ricardo Gherardi Hein

Received: October 10/2024
Accepted: May 6/2025

INTRODUCTION

The Caatinga is a Brazilian biome that has approximately 3 347 species, 962 genera and 153 families of plants, 43.7% of them classified as woody. The principal families are Fabaceae, Myrtaceae, Euphorbiaceae, Rutaceae, Malvaceae, Rubiaceae, Bignoniaceae and Sapindaceae, while the main genera are *Croton* L., *Mimosa* L., *Chamaecrista* (L.) Moench, *Senna* Mill. and *Eugenia* L. (Fernandes *et al.*, 2020). These occur in the states of Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia and the northern region of Minas Gerais. The biome's sustainable use and forest resources conservation are important for commerce of wood and non-wood products (Pareyn *et al.*, 2024).

In 2023, Caatinga biome presented a deforestation of 11% from its area, representing 201 687 ha, being the third biome degraded in this year, with increase of 43% related to 2022. Cerrado was the first, with 1 110 326 ha, and Amazon the second with 454 271 ha (MapBiomas, 2024). The biome has a high deforestation index due to increasing population density, conversion of forest areas to farmland and grazing pasture, and aggressive exploitation of natural resources, resulting in climate change and sometimes desertification (Demartelaere *et al.*, 2021).

The caatinga has a great diversity of cultural applications, although it is composed of small trees. Native Caatinga tree species can legally be cut for firewood and other uses through authorization, demonstrated by a "Forest Origin Document" (DOF), indicating that the wood is from an area with an active forest management plan. Illegal exploitation of wood aggravates deforestation in tropical forests, and many species are listed as at risk according to the Convention on International Trade in Endangered Species of Wild Flora and Fauna (CITES). An alternative for adequate use of forest resources is to limit exploitation to areas under certified sustainable management, involving replanting with native species, because many of them are endemic to the Caatinga. However, for this system to function properly, correct identification of species based on taxonomic information is necessary, so as to deter illegal logging (Moraes *et al.*, 2018).

In Caatinga, not always leaves are available for analysis once many species are deciduous tree, i.e., loose leaf in some climatic period in the year. So, it is necessary the use of other tree part for identification. Bark can be an alternative in function of its availability and easiness of access through the seasons in forest, and also, thinking in commerce, in storage of logs in industry.

It is known that the morphological and macroscopic characteristics of bark can be applied for this identification, but alone these are not very effective due to the natural variability and similar general texture pattern of various species. Other drawbacks are the lengthy testing process and subjectivity of human optical sensitivity (Kim *et al.*, 2022; MacFarlane, 2024).

Near-infrared spectroscopy (NIR) can be an alternative because it is fast and the analysis can be accomplished with a great range of material, such as leaves, bark and wood (Zhou *et al.*, 2020; Tsuchikawa *et al.*, 2023; Silva *et al.*, 2021, Silva *et al.*, 2024). Regarding Caatinga species, Nisgoski *et al.* (2018)

evaluating wood and charcoal to verify the potential of NIR to discriminate six species in the municipality of Coremas, Paraíba state, described better results applying principal component analysis with linear discriminant analysis (PCA-LDA) and the use of second derivative spectra, highlighting there was no influence of anatomical surface in material distinction, which is important for practical applications.

With regard to bark spectra, studies have evaluated the chemical composition (Acquah *et al.*, 2015; Bridson *et al.*, 2024), presence of bark in waste composition (Acquah *et al.*, 2016), influence of cutting/analysis period (Kim *et al.*, 2013), and quantification of salicylates and flavonoids in poplar bark (Mazurek *et al.*, 2022). The separation of *Quercus suber* L. trees from two geographic regions for analysis of the quality of cork planks (Prades *et al.*, 2010), and the discrimination of seven species from Costa Rica, with accuracy of 74.2% (Clark and Roberts, 2012), are among the first works involving distinction of trees based on bark spectra.

Using bark spectra obtained with portable spectrometer, Hadlich *et al.* (2018) evaluated 254 trees (eleven species, ten genera, eight families) with Principal Component Analysis and Linear Discriminant Analysis, and obtained correct identification of 98% (inner bark) and 94% (outer bark). Other authors discriminated *Betula pendula*, *Pinus sylvestris* and *Picea abies* by applying Support Vector Machine and a mobile handheld hyperspectral camera (Juola *et al.*, 2020), and reiterated intra/interspecific variation of bark reflectance of boreal/temperate tree species (Juola *et al.*, 2022).

There are some limitations on the use of NIR for species identification in practice. Chiefly, accuracy is dependent on the quality of the calibration model and spectral variability between and within species, which also can be affected by surface contamination, moisture content and texture. Also, a large database is required to ensure that spectral data are really representative of the species being evaluated (Tsuchikawa *et al.*, 2023).

Considering the importance of species identification and the scarcity of literature relating NIR spectroscopic analysis to Caatinga species, we aimed to discriminate the species: *Anadenanthera colubrina*, *Cenostigma pyramidale*, *Capparidastrium frondosum*, *Commiphora leptophloeos*, *Mimosa tenuiflora*, *Manihot baccata*, *Guapira* sp. and *Aspidosperma pyrifolium* from the northeastern region of Brazil based on bark analysis, increasing research database, for application in forestry supervision.

MATERIALS AND METHODS

The farm where species were cut has an active sustainable forest management plan for exploitation of wood for energy generation. Species (Tab. 1) were selected based on frequency and index of importance values, i.e. based on number of trees in each species and its dimensions, in function of forest inventory to development of the management plan. To develop a robust model and to simulate the obtention of spectra in industry when the position is unknown, the trunks were divided into six positions: 0% (10 cm from ground), diameter at breast height (DBH - 1.30 m from ground), 25%, 50%, 75% and 100% of commercial height. A schematic representation of methodology is in figure 1.

Material

The Caatinga species (Tab. 1, Fig. 2) were collected from the Riacho do Cabra farm, in the municipality of Santa Cruz, Rio Grande do Norte, Brazil (06°13'44.4"S, 36°01'22.8"W). The climate is classified as BSs'h' on the

Köppen scale, semiarid with precipitation lower than evaporation, a rainy season in autumn and dry season lasting six months. In the period of 2022-2024 mean annual precipitation varied from 581 mm to 815 mm, mean annual temperature from 26 °C to 28 °C and mean annual relative humidity is approximately 70% (EMPARN, 2024).

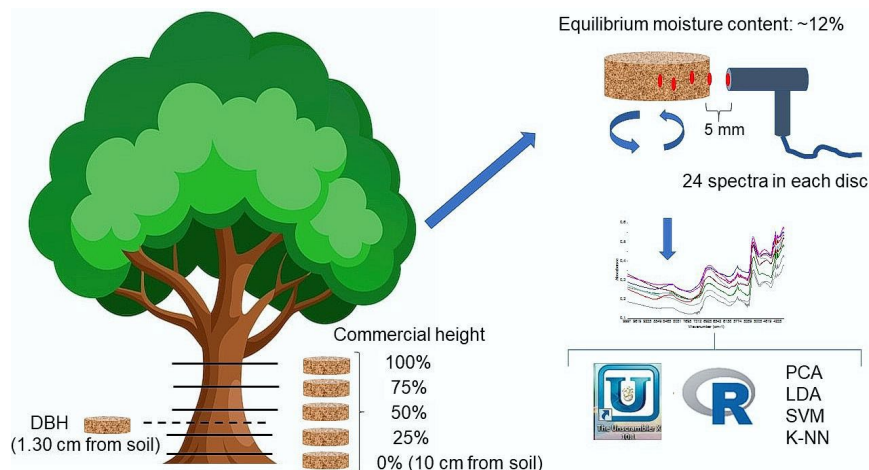


Figure 1: Schematic representation of cut position, spectra obtention and analysis for evaluated caatinga species.

Table 1: Information about the vernacular and scientific names, code, diameter at breast height, commercial and total height of the individuals analyzed.

Vernacular name in Portuguese	Code	Scientific name - Family	DBH (cm)	Commercial Height (m)	Total height (m)
Angico	ANG	<i>Anadenanthera colubrina</i> (Vell.) Brenan - Fabaceae	14.01	6.7	10.00
			15.92	4.0	11.20
			22.12	5.0	10.96
Catingueira	CAT	<i>Cenostigma pyramidale</i> (Tul.) E. Gagnon & G.P. Lewis - Fabaceae	14.64	4.6	7.86
			14.96	3.5	8.43
			14.01	4.0	8.18
Feijão Bravo	FEB	<i>Capparidastrium frondosum</i> (Jacq.) Cornejo & Iltis - Capparaceae	14.01	5.0	6.92
			13.37	4.4	6.90
			13.32	3.3	8.39
Imburana	IMB	<i>Commiphora leptophloeos</i> (Mart) J.B. Gillett - Burseraceae	12.73	2.7	6.26
			15.28	3.5	8.37
			23.55	4.5	8.26
Jurema Preta	JUP	<i>Mimosa tenuiflora</i> (Willd.) Poir. - Fabaceae	14.96	4.1	9.15
			17.19	2.9	6.52
			16.55	3.9	7.13
Maniçoba	MAN	<i>Manihot baccata</i> Allem - Euphorbiaceae	12.73	4.5	8.20
			14.01	5.7	8.60
			20.05	5.3	8.56
Pau Mole	PAM	<i>Guapira</i> sp. - Nyctaginaceae	18.78	5.5	7.73
			15.28	4.0	8.30
			13.37	3.05	6.77
Pereiro	PER	<i>Aspidosperma pyrifolium</i> Mart. & Zucc. - Apocynaceae	12.73	3.05	6.73
			13.37	2.67	6.45
			13.37	2.85	7.0

DBH=Diameter at breast height.

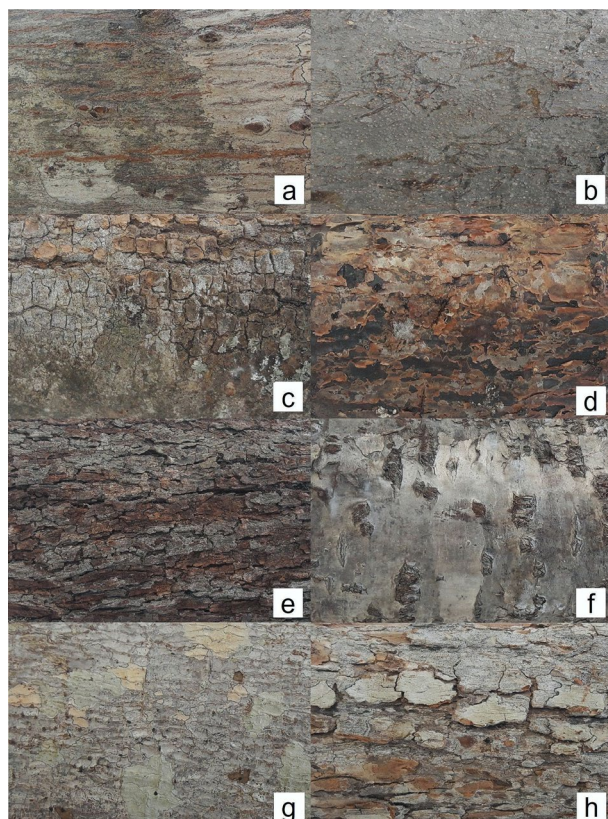


Figure 2: Illustration of tree bark in DBH from studied species: ANG (a), CAT (b), FEB (c), IMB (d), JUP (e), MAN (f), PAM (g) e PER (h).

Three trees were cut from species and had eight years old when cut. The vegetation in the region is classified as hyperxerophyllous Caatinga. The access to the material is registered with the Brazilian Council for Management of Genetic Heritage (CGEN/SISGEN) under number A432EB4.

Near infrared spectroscopy

All discs remained at the same ambient temperature until reaching an equilibrium moisture content with the environment of approximately 16%, and before analysis the samples were stored in a climate-controlled room with temperature of 25 ± 2 °C and relative humidity of 60%, to reach approximately 12% moisture content.

For spectral analysis, a Bruker Tensor 37 spectrometer (Bruker Optics, Ettlingen, Germany) equipped with a probe with 2 mm aperture was used. Spectra were collected with resolution of 4 cm^{-1} , 64 scans and wavenumber range from $10\,000$ to $4\,000 \text{ cm}^{-1}$. In function of irregularities in bark surface, all discs from the six trunk positions were put at approximately 5 mm from the probe, and 24 spectra were collected from each disc, for a total of 144 per tree, 432 per species, totaling 3 456 for analysis. Region where spectra was collected varied to better represent structure and composition of bark.

Multivariate Analysis

Data were analyzed in raw form and after first- and second-derivative processing of Savitzky-Golay, both in polynomial order = 2 and smoothing points = 21. Exploratory modeling was done by visual analysis of the score and loading graphs obtained by principal component analysis (PCA), to verify possible differences/groupings of species. PCA was performed based on the NIPALS algorithm, with random cross-validation using 3 segments.

To verify the possibility of external discrimination of species, three classifications were performed: linear discriminant analysis (LDA), k-NN and Support Vector Machine (SVM) analysis. Spectra were divided into 66% for model construction and 33% for external analysis, i.e., data from two trees for model construction and one tree for species classification, in two situations: i) using all tree positions; and ii) using only the DBH position. LDA was performed with PCA scores based on the Mahalanobis distance, projecting 3 components assuming equal prior probabilities. Support vector machines with kernel radial basis function (SVM) and k-nearest neighbors (k-NN) were tested using the function “train” for different methods available in the “caret” package in the R software (Kuhn *et al.*, 2020). External classification was evaluated with all wavenumbers in raw form and after first- and second-derivative transformation. PCA and LDA were performed with the Unscrambler X chemometric program (version 10.1, CAMO Software, Oslo, Norway, www.camo.com).

RESULTS

Mean bark spectra by species (Fig. 3) indicated the materials similarity, and some regions showed good information for differentiation, for example, wavenumbers from $8\,470 - 8\,150 \text{ cm}^{-1}$ and from $4\,930 - 4\,490 \text{ cm}^{-1}$.

To verify visual grouping, principal component analysis (PCA) were performed with spectra in raw form and after first- and second-derivative transformation (Fig. 4). PCA with raw spectra (Fig. 4a) identified a mixture of all species, although some fine distinction of MAN and PER samples were observed. PCA with first-derivative spectra (Fig. 4b) indicated more proximity of samples inside each species and their better identification, but PC1 represented only 35% of variations between species. According to PCA with second-derivative spectra (Fig. 4c), the first PC represented 64% of the spectral variation between species, and the samples from CAT, FEB and PER were divided into two groups, probably in function of intrinsic bark variation based on tree height.

To verify the potential of NIR spectra from tree bark to differentiate the Caatinga species, LDA, k-NN and SVM classification were tested with raw data along with first- and second-derivative spectra by applying all tree positions (Tab. 2) and only with DBH position (Tab. 3). Table 4 indicates correct percentage of classification for each species, pretreatment and model applied and the accuracy of all performed tests is reported in Table 5. The results reinforce the influence of pretreatment and classification methods for analysis.

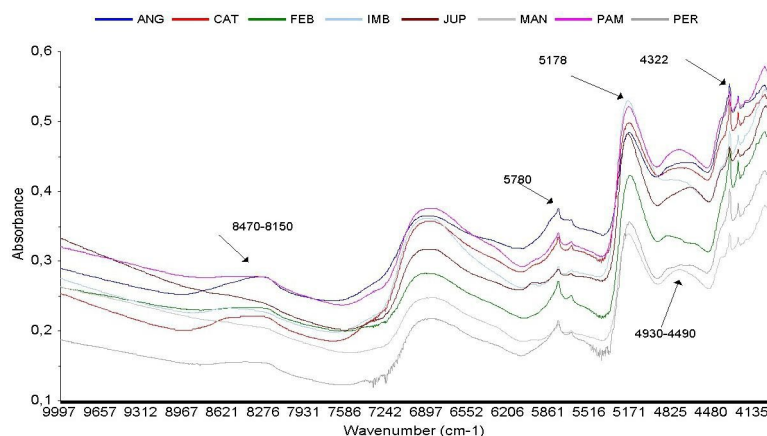


Figure 3: Mean NIR raw spectra from all positions of bark samples of eight Caatinga species.

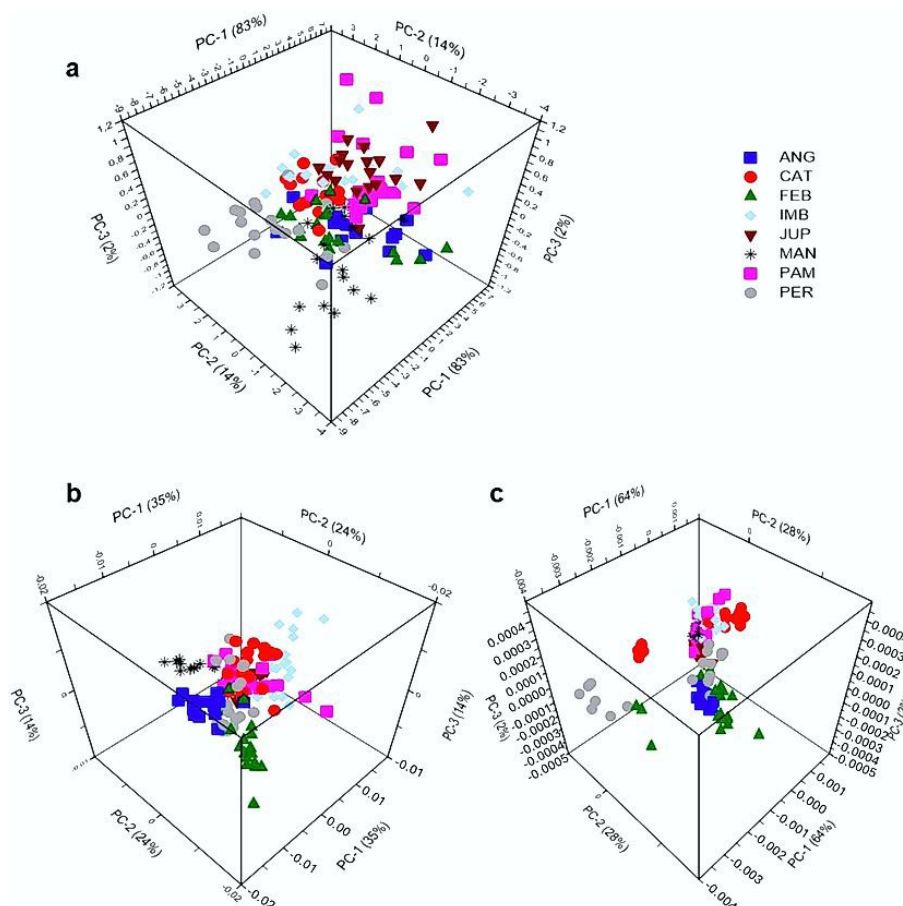


Figure 4: Graphic scores of principal component analysis of raw spectra (a), first derivative (b) and second derivative (c) spectra.

DISCUSSION

In bark tree spectra (Fig. 3), the signals in the region from 8 470 – 8 150 cm^{-1} were attributed to cellulose, hemicellulose and lignin; a band at 5 795 cm^{-1} was related to lignin, one at 5 776 cm^{-1} to cellulose, and another at 5 178

cm^{-1} to water. Wavenumbers at 4 930 and 4 490 cm^{-1} were attributed to cellulose, hemicellulose, lignin and extractives (Schwanninger *et al.*, 2011).

Bark NIR spectra can be influenced by genetic factors and tree provenance, as sometimes can occur species natural hybridization and/or adaptation in diverse

habitats, for example, to improve resistances against drought and UV-radiation, increasing growing rates and altering its properties (Zhang et al., 2014), as well as by the surface evaluated (Prades et al., 2010; Hadlich et al., 2018).

Bark in general absorbs less light than leaves (Acquah et al., 2015) and its spectral properties can vary in the same tree and/or between trees, as well as in function of water and chemical content (Clark and Roberts, 2012). The chemical composition and thickness of bark has a direct relation with herbivory, due to chemical defense of the plant, and environmental stressors, as fire, frost, drought and waterlogging (Ilek et al., 2021; MacFarlane, 2024). It is more variable than wood and has high content of lignin, extractives and inorganic material, with lower percentage of polysaccharides (Heim et al., 2022; Supriyadi et al., 2025).

Score graphic of PCA (Fig. 4) indicate grouping formation of samples from the same tree after derivative pretreatments. Some species can be distinguished from others, what is corroborated by visual observation of the bark surface (Fig. 2), which indicated irregular surfaces with thin bark, except for JUP, and flaked appearance for IMB and MAN. Variable bark thickness can be used to

distinguish species, but also can vary between trees of the same species due to herbivory, fire, biomechanical support, respiration, age, stem position and climatic conditions (Shearman and Varner, 2021; Nie et al., 2022).

In this study, we evaluated young trees and observed that spectra from bark in different axial trunk position had some differences and influence in correct species discrimination. The analysis of oldest trees must be tested and compared, but some standardization on position of spectra obtention at DBH can contribute to practical applications, as the factor *species* is the most important, and properties are more variable and important in lower positions of tree (Rodriguez-Perez et al., 2022).

The use of spectra only from the DBH position increased correct classification of species (Tab. 4), as did the use of derivatives. LDA had the lowest correct classification. K-NN and SVM performed well using 1st and 2nd derivative, varying in function of species. For k-NN, five species were correctly classified above 91% with first derivative and 6 species with second derivative, while according to SVM, one species was adequately identified with first derivative and 5 with second derivative.

Table 2: Confusion matrix of external classification by LDA, KNN and SVM based on NIR spectra in all tree positions (n=144 in each species).

LDA – Raw data									KNN – Raw data									SVM – Raw data								
	ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER
ANG	82	15	0	8	1	3	3	3	ANG	85	0	2	2	1	0	2	12	ANG	100	3	0	1	0	0	3	19
CAT	4	4	0	2	0	0	0	0	CAT	9	96	1	17	6	5	21	6	CAT	4	110	2	11	4	5	10	2
FEB	10	119	59	13	12	15	12	11	FEB	19	5	94	5	5	5	13	27	FEB	9	3	104	4	1	2	14	26
IMB	37	3	22	74	38	4	106	26	IMB	0	7	11	73	0	0	6	4	IMB	10	4	12	107	4	13	20	3
JUP	0	0	4	2	53	1	4	2	JUP	5	6	1	15	119	8	9	1	JUP	2	3	0	5	127	4	2	3
MAN	0	2	22	7	12	114	12	37	MAN	1	3	4	13	11	112	20	22	MAN	0	4	2	11	5	102	17	14
PAM	9	0	8	8	4	5	7	2	PAM	24	25	18	17	1	5	67	14	PAM	15	14	12	2	2	14	73	9
PER	2	0	29	30	24	2	0	63	PER	1	2	13	2	1	9	6	58	PER	4	3	12	3	1	4	5	68
LDA – First derivative									KNN – First derivative									SVM – First derivative								
	ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER
ANG	97	107	5	5	19	30	13	48	ANG	130	0	0	0	0	0	0	5	ANG	3	0	0	0	0	0	0	0
CAT	5	29	0	15	1	3	31	0	CAT	0	144	0	3	0	0	2	1	CAT	0	10	0	0	0	0	0	0
FEB	0	4	82	0	0	0	1	17	FEB	1	0	125	0	1	0	5	14	FEB	0	0	34	0	0	0	0	1
IMB	5	0	2	109	8	0	82	4	IMB	0	0	1	132	0	0	4	0	IMB	141	134	109	144	75	92	126	109
JUP	1	0	8	6	100	0	1	2	JUP	4	0	2	3	143	0	0	4	JUP	0	0	0	0	69	0	0	0
MAN	0	0	7	1	5	101	1	2	MAN	0	0	0	6	0	140	6	0	MAN	0	0	0	0	0	52	0	0
PAM	5	0	4	7	9	10	9	4	PAM	3	0	12	0	0	4	123	5	PAM	0	0	1	0	0	0	18	0
PER	31	4	36	1	2	0	6	67	PER	6	0	4	0	0	0	4	115	PER	0	0	0	0	0	0	0	34
LDA – Second derivative									KNN – Second derivative									SVM – Second derivative								
	ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER
ANG	17	0	0	0	0	7	0	18	ANG	110	0	0	0	0	0	0	3	ANG	0	0	0	0	0	0	0	0
CAT	43	0	0	50	0	38	8	39	CAT	25	144	0	0	0	0	0	0	CAT	0	0	0	0	0	0	0	0
FEB	71	2	128	58	34	3	65	74	FEB	0	0	141	0	2	0	0	2	FEB	0	0	0	0	0	0	0	0
IMB	6	0	3	0	1	18	18	4	IMB	0	0	0	144	2	0	0	0	IMB	3	0	1	144	1	0	0	1
JUP	0	0	3	14	100	1	7	0	JUP	0	0	0	0	91	0	0	0	JUP	19	0	20	0	69	0	0	0
MAN	0	0	0	15	3	50	38	0	MAN	0	0	0	0	48	24	0	0	MAN	3	0	1	0	72	24	0	1
PAM	3	0	10	6	6	27	5	7	PAM	6	0	2	0	0	120	128	0	PAM	28	0	24	0	0	120	127	0
PER	4	142	0	1	0	0	1	2	PER	3	0	1	0	1	0	16	139	PER	91	144	98	0	2	0	17	142

Table 3: Confusion matrix of external classification by LDA, KNN and SVM based on NIR spectra in DBH position (n=24 for each species).

LDA – Raw data									KNN – Raw data									SVM – Raw data								
	ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER
ANG	9	0	0	0	0	0	1	0	ANG	13	0	0	0	0	0	3	0	ANG	9	0	0	0	0	0	0	0
CAT	1	2	0	0	0	0	0	0	CAT	0	14	0	5	0	0	6	1	CAT	0	13	0	2	0	0	6	0
FEB	1	0	9	0	0	1	2	1	FEB	5	0	14	1	0	0	2	9	FEB	3	1	16	1	0	0	1	5
IMB	7	18	6	21	20	1	19	2	IMB	1	3	1	5	0	0	2	1	IMB	9	3	2	12	0	11	6	2
JUP	0	0	0	1	2	0	0	0	JUP	0	0	0	4	22	1	2	1	JUP	0	0	0	0	22	1	0	0
MAN	3	2	1	0	1	21	1	6	MAN	1	1	1	1	1	9	3	3	MAN	1	5	1	4	2	12	4	2
PAM	2	2	1	0	1	1	1	0	PAM	4	3	3	5	1	3	6	0	PAM	0	2	2	3	0	0	7	2
PER	1	0	7	2	0	0	0	15	PER	0	3	5	3	0	11	0	9	PER	2	0	3	2	0	0	0	13

LDA – First derivative									KNN – First derivative									SVM – First derivative								
	ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER
ANG	15	0	0	0	0	0	0	2	ANG	22	0	0	0	0	0	0	0	ANG	0	0	0	0	0	0	0	0
CAT	2	20	0	0	2	0	0	0	CAT	0	24	0	2	0	0	2	0	CAT	0	9	0	0	0	0	0	0
FEB	1	0	18	0	0	0	0	2	FEB	0	0	23	0	0	0	0	7	FEB	0	0	8	0	0	0	0	1
IMB	0	0	0	21	1	0	21	1	IMB	0	0	0	22	0	0	2	0	IMB	24	15	16	24	20	23	24	22
JUP	0	0	0	0	12	0	0	0	JUP	0	0	0	0	24	0	0	0	JUP	0	0	0	0	4	0	0	0
MAN	0	0	0	0	1	17	0	0	MAN	0	0	0	0	0	13	0	0	MAN	0	0	0	0	0	1	0	0
PAM	1	0	1	3	8	6	0	3	PAM	0	0	0	0	0	11	20	1	PAM	0	0	0	0	0	0	0	0
PER	5	4	5	0	0	1	3	16	PER	2	0	1	0	0	0	0	16	PER	0	0	0	0	0	0	0	1

LDA – Second derivative									KNN – Second derivative									SVM – Second derivative								
	ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER
ANG	6	0	0	0	0	0	1	1	ANG	22	0	0	0	0	0	0	0	ANG	0	0	0	0	0	0	0	0
CAT	12	7	3	3	3	0	0	7	CAT	0	24	0	0	0	0	0	1	CAT	0	24	0	0	0	0	0	1
FEB	2	0	3	0	0	0	9	5	FEB	1	0	24	0	1	0	0	7	FEB	0	0	24	0	1	0	0	1
IMB	0	0	4	0	0	0	10	1	IMB	0	0	0	24	0	0	0	0	IMB	0	0	0	24	0	0	2	0
JUP	0	0	0	0	11	0	0	1	JUP	0	0	0	0	23	0	0	0	JUP	0	0	0	0	23	0	0	0
MAN	0	0	0	7	0	0	0	0	MAN	0	0	0	0	0	0	0	0	MAN	0	0	0	0	0	0	0	0
PAM	0	0	3	9	10	11	0	0	PAM	0	0	0	0	0	24	22	0	PAM	0	0	0	0	0	24	21	0
PER	4	17	11	5	0	13	4	9	PER	1	0	0	0	0	0	2	16	PER	24	0	0	0	0	0	1	22

There was no pattern of bark characteristics for incorrect classification of samples, which can be explained in function of morphological characteristics of bark related to different surface irregularities, or perhaps the presence of fungi. This variation is in accordance with the results described in Amazonian trees, where the authors reported the absence of a clear pattern for correct or incorrect classification (Hadlich et al., 2018), and analysis in a boreal forest (Juola et al., 2022).

Other authors have also described that SVM produces better results than LDA, such as for discrimination of western hemlock and hem-fir green mix of timber (Zhou et al., 2020). SVM was also efficient in discrimination of *Betula pendula*, *Pinus sylvestris* and *Picea abies* stem bark (Juola et al., 2020) and was described as the best classification method for determining biofuel quality based on NIR spectra (Mancini et al., 2019). Similarly, SVM had an accuracy of 93% in discriminating Chinese trees with different origins (Li et al., 2022), and accuracy above 99% in identification of 25 wood species applied in floors (Pan et al., 2021). On the other hand,

Sem et al. (2018) reported that LDA was more sensitive to class-imbalance than SVM.

Table 5 indicates that better accuracy results were obtained with K-NN and first-derivative spectra. Lower accuracy (spectra in all tree positions and from DBH) were obtained for LDA classification by applying spectra after second-derivative adjust. It is important to comment that other author's obtained good results using LDA to distinguish Amazonian species with bark samples (Hadlich et al., 2018), highlighting the importance of studies with multivariate analysis and different plant tissues and also different biomes.

Although bark typically has more intraspecific spectral variation, as a result of environmental effects, thickness and plant age, which can hinder the discrimination power of NIR (Hadlich et al., 2018), our results indicate the technique's potential to identify Caatinga species based on bark spectra. It is important to evaluate the same species with different ages in forest and after cut, as logs, before debarking in industry, to confirm the accuracy of technique in practical applications.

Table 4: Correct classification in external prediction for Caatinga species discrimination.

All tree positions									
Method/transformation		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER
LDA	Raw spectra	56.9	2.7	41.0	51.4	36.8	79.2	4.9	43.7
	1 st derivative	67.4	20.1	56.9	75.7	69.4	70.1	6.2	46.5
	2 nd derivative	11.8	0	88.9	0	69.4	34.7	3.5	1.4
k-NN	Raw spectra	59.0	66.7	65.3	50.7	82.6	77.8	46.5	40.3
	1 st derivative	90.3	100	86.8	91.7	99.3	97.2	85.4	79.9
	2 nd derivative	76.4	100	97.9	100	63.2	16.7	88.9	96.5
SVM	Raw spectra	69.4	76.4	72.2	74.3	88.2	70.8	50.7	47.2
	1 st derivative	2.1	6.9	23.6	100	47.9	36.1	12.5	23.6
	2 nd derivative	0	0	0	100	47.9	16.6	88.2	98.6
Only DBH position									
Method		ANG	CAT	FEB	IMB	JUP	MAN	PAM	PER
LDA	Raw spectra	37.5	8.3	37.5	87.5	8.3	87.5	4.2	62.5
	1 st derivative	62.5	83.3	75.0	87.5	50.0	70.8	0	66.7
	2 nd derivative	25.0	29.2	12.5	0	45.8	0	0	37.5
k-NN	Raw spectra	54.2	58.3	58.3	20.8	91.7	37.5	25.0	37.5
	1 st derivative	91.7	100	95.8	91.7	100	54.2	83.3	66.7
	2 nd derivative	91.7	100	100	100	95.8	0	91.7	66.7
SVM	Raw spectra	37.5	54.2	66.7	50.0	91.7	50.0	29.2	54.2
	1 st derivative	0	37.5	33.3	100	16.7	4.2	0	4.2
	2 nd derivative	0	100	100	100	95.8	0	87.5	91.7

Table 5: Accuracy of tests performed to distinguish Caatinga species.

Analysis	Transformation	Accuracy (%)	
		Spectra in all tree position	Spectra from DBH
LDA	Raw data	0.42	0.42
	1 st derivative	0.52	0.62
	2 nd derivative	0.26	0.19
SVM	Raw data	0.69	0.54
	1 st derivative	0.32	0.25
	2 nd derivative	0.44	0.72
K-NN	Raw data	0.61	0.48
	1 st derivative	0.91	0.85
	2 nd derivative	0.81	0.81

Also, the study of other species from the same region or other forest origin, with variations in point of spectra obtention to confirm that the DBH is the position more indicate to distinction of material. The use of near infrared in forest can confirm the correct species before cut on forest management, contributing to conservation of Caatinga resources and an adequate use of species with high aggregated value.

CONCLUSIONS

The accuracy values ranged from 0.19 to 0.91, varying in function of different pretreatments of spectral data and also classification methods. Near-infrared spectroscopy with multivariate analysis has potential to discriminate Caatinga

species. Considering the diverse discriminant models tested, better results were obtained for K-NN classification with first-derivative spectra, with data obtained at all tree height positions. Also, satisfactory results were obtained only applying spectra obtained in DBH position, making the access and spectra obtention easiness in forest. It is important to note that for application in the field, more data on these species are necessary.

ACKNOWLEDGMENTS

Authors thanks the owner of "Fazenda Riacho do Cabra", in Santa Cruz, Rio Grande do Norte, for material donation, National Council for Scientific and

Technological Development (CNPQ–303352/2022-1), the Araucaria Foundation (project 4206), and Office to Coordinate Improvement of University Personnel (CAPES, Finance Code 001).

AUTHORSHIP CONTRIBUTION

Project Idea: SN, SHBG, RCS, GIBM

Funding: SN; GIBM

Database: SN, HCV, JXS, SHBG, GIBM

Processing: SN, HCV, JXS, SHBG, RCS, GIBM

Analysis: HCV, JXS, SHBG

Writing: SN, HCV, JXS, SHBG, GIBM

Review: SN, HCV, JXS, SHBG, RCS, GIBM

REFERENCES

- ACQUAH, G.E.; VIA, B.K.; FASINA, O.O.; et al. Non-destructive prediction of the properties of forest biomass for chemical and bioenergy applications using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*. v. 23, n. 2, p. 93-102, 2015. <https://opg.optica.org/jnirs/abstract.cfm?URL=jnirs-23-2-93>
- ACQUAH, G.E.; VIA, B.K.; BILLOR, N.; et al. Identifying plant part composition of forest logging residue using infrared spectral data and linear discriminant analysis. *Sensors*. v.16, n.9, 1375, 2016. <https://doi.org/10.3390/s16091375>
- BRIDSON, J.H.; THUMM, A.; COOKE-WILLIS, M. et al. Comparison of near infrared and mid infrared spectroscopy for the prediction of *Pinus radiata* bark chemical properties relevant to biorefinery. *Biomass and Bioenergy*. v.185, 107235, 2024. <https://doi.org/10.1016/j.biombioe.2024.107235>
- CLARK, M.L.; ROBERTS, D.A. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sensing*. v.4, n. 6, p.1820–1855, 2012. <https://doi.org/10.3390/rs4061820>
- DEMARTELAERE, A.C.F.; SANTOS, A.F.; COUTINHO, P.W.R. et al. Causas, consequências, e métodos atribuídos para prevenir a desertificação na Caatinga. *Brazilian Journal of Development*, v.7, n.8, p.83270-83285, 2021. <https://doi.org/10.34117/bjdv7n8-502>
- EMPARN (Empresa de Pesquisa Agropecuária do Rio Grande do Norte). Posto: EMATER (Santa Cruz) – 2024. Natal, RN: EMPARN, 2024. <https://meteorologia.emparn.rn.gov.br/relatorios/relatoriospluviometricos?tipo=acumuladaCorrente>.
- FERNANDES, M.F.; CARDOSO, D.; QUEIROZ, L.P. An updated plant checklist of the Brazilian Caatinga seasonally dry forests and woodlands reveals high species richness and endemism. *Journal of Arid Environments*. v.174, p.104079, 2020. <https://doi.org/10.1016/j.jaridenv.2019.104079>
- HADLICH, H.L.; DURGANTE, F.M.; SANTOS, J.D.; et al. Recognizing Amazonian tree species in the field using bark tissues spectra. *Forest Ecology and Management*. v.427, p. 296–304, 2018. <https://doi.org/10.1016/j.foreco.2018.06.002>
- HEIM, L.; DODELER, R.; BRANCHERIAU, L. et al. Comparison of the extractives chemical signatures among branch, knot and bark wood fractions from forestry and agroforestry Walnut (*Juglans regia* × *J. nigra*), by NIR-Spectroscopy and LC-MS analyses. *iForest*. V.15, n.1, p. 56-62, 2022. <https://doi.org/10.3832/ifor3973-014>
- ILEK, A.; SIEGERT, C.M.; WADE, A. Hygroscopic contributors to bark storage and controls exerted by internal bark structure over water vapor absorption. *Trees*. v. 35, p.831-843, 2021. <https://doi.org/10.1007/s00468-021-02084-0>
- JUOLA, J.; HOVI, A.; RAUTIAINEN, M. Multiangular spectra of tree bark for common boreal tree species in Europe. *Silva Fennica*. v.54, n.4, p. 1-18, 2020. <https://doi.org/10.14214/sf.10331>
- JUOLA, J.; HOVI, A.; RAUTIAINEN, M. A spectral analysis of stem bark for boreal and temperate tree species. *Ecology and Evolution*. v.12, n. 3, p. 18 – 87, 2022. <https://doi.org/10.1002/ece3.8718>
- KIM, N.; YU, M.; LEE, D.Y. et al. Rapid determination of betulin in *Betula platyphylla* outer bark using near-infrared spectroscopy. *Analytical Letters*. v.46, p. 1289-1298, 2013. <https://doi.org/10.1080/00032719.2012.755686>
- KIM, T.K.; HONG, J.; RYU, D. et al. Identifying and extracting bark key features of 42 tree species using convolutional neural networks and class activation mapping. *Scientific Reports*. n.12, p.4772, 2022. <https://doi.org/10.1038/s41598-022-08571-9>
- KUHN, M.; WING, J.; WESTON, S.; et al. Package *caret*. *The R Journal*, 2020. Disponível em: <https://CRAN.R-project.org/package=caret>.
- LI, Y.; VIA, B.K.; LI, Y. WANG, G. Determination of geographical origin and tree species using VIS-NIR and chemometric methods. *Forest Products Journal*. v.72, p. 147–154, 2022. <https://www.doi.org/10.13073/FJ-P-D-22-00011>
- MACFARLANE, D.W. Highly variable bark-wood density relationships across tree species reflect tradeoffs in evolved tolerances to environmental stressors. *Trees*. v.38, p.1223–1239, 2024. <https://doi.org/10.1007/s00468-024-02548-z>
- MANCINI, M.; TAAVITSAINEN, V.; TOSCANO, G. Comparison of three different classification methods performance for the determination of biofuel quality by means of NIR spectroscopy. *Journal of Chemometrics*. v. 33, n. 7, p. 31-45, 2019. <https://doi.org/10.1002/cem.3145>
- MAPBIOMAS. RAD2023 - Relatório anual do desmatamento no Brasil 2023. São Paulo, Brasil: MapBiomas, 2024. 154p. https://alerta.mapbiomas.org/wp-content/uploads/sites/17/2024/10/RAD2023_COMPLETO_15-10-24_PORTUGUES.pdf
- MAZUREK, S.; WŁODARCZYK, M.; PIELORZ, S.; et al. Quantification of salicylates and flavonoids in poplar bark and leaves based on IR, NIR, and raman spectra. *Molecules*. V.27, n. 12, p. 3954, 2022. <https://doi.org/10.3390/molecules27123954>
- MORAIS, R.M.; CUNHA, M.D.C.L.; SANTANA, G.M.; et al. Dendrological characterization as inspection resources of caatinga wood market. *Floresta e Ambiente*. v.25, n. 3, e20170813, 2018. <https://doi.org/10.1590/2179-8087.081317>
- NIE, W.; LIU, Y.; TAN, C. et al. Characteristics and factors driving the variations in bark thickness of major woody plants in China. *Ecological Indicators*. n.144, 109447, 2022. <https://doi.org/10.1016/j.ecolind.2022.109447>
- NISGOSKI, S.; BATISTA, F.R.R.; NAIDE, T.L.; et al. Discrimination of wood and charcoal from six Caatinga species by near-infrared spectroscopy. *Maderas. Ciencia y Tecnología*. v. 20, n. 2, p. 199-210, 2018. <http://dx.doi.org/10.4067/S0718-221X2018005002401>
- PAN, X.; LI, K.; CHEN, Z.; YANG, Z. Identifying wood based on near-infrared spectra and four gray-level co-occurrence matrix texture features. *Forests*. v.12, n.11, 1527, 2021.
- PAREYN, F.G.C.; RIEGELHAUPT, E.M.; GARLET, A.; et al. Manejo florestal na Caatinga: 40 anos de experimentação. Brasília, DF: Ministério do Meio Ambiente, 2024. 458p.
- PRADES, C.; GARCÍA-OLMO, J.; ROMERO-PRIETO, T.; et al. Methodology for cork plank characterization (*Quercus suber* L.) by near-infrared spectroscopy and image analysis. *Measurement Science & Technology*. v. 21, n.6, p.065- 602, 2010. DOI 10.1088/0957-0233/21/6/065602
- RODRIGUEZ-PEREZ, D.; MOYA, R.; MURILLO, O. Effect of stem height in variation of bark, heartwood, sapwood and physical properties of wood in *Dipteryx panamensis* Pittier in a provenance/progeny test. *Ciência Florestal*. v. 32, n. 1, p. 141-162, 2022. <https://doi.org/10.5902/1980509843606>
- SCHWANNINGER, M.; RODRIGUES, J.C.; FACKLER, K. A review of band assignments in near infrared spectra of wood and wood components. *Journal of Near Infrared Spectroscopy*. v.19, n. 5, p. 287-308, 2011. <https://doi.org/10.1255/jnirs.955>

SEM, V.; KOLAR, J.; LUSA, L. Artificially generated near-infrared spectral data for classification purposes. *Chemometrics and Intelligent Laboratory Systems*. v.172, p.100-108, 2018. <https://doi.org/10.1016/j.chemolab.2017.11.009>

SHEARMAN, T.M.; VARNER, J.M. Variation in bark allocation and rugosity across seven co-occurring southeastern US tree species. *Frontiers in Forest and Global Change*. v.4, 731020, 2021. <https://doi.org/10.3389/ffgc.2021.731020>

SILVA, A.L.D.; ALVES FILHO, E.G.; SILVA, L.M.A. et al. Near infrared spectroscopy to rapid assess the rubber tree clone and the influence of maturation and disease at the leaves. *Microchemical Journal*. v.168, 106478, 2021. <https://doi.org/10.1016/j.microc.2021.106478>

SILVA, C.E.; NASCIMENTO, C.S.; FREITAS, J.A. et al. Alternative identification of wood from natural fallen trees of the Lecythidaceae family in the Central Amazonian using FT-NIR spectroscopy. *International Forestry Review*. v.26, n.1, p.29-44, 2024. <https://doi.org/10.1505/146554824838457844>

SUPRIYADI, D.; DAMAYANDI, D.; VEIGEL, S. et al. Unlocking the potential of tree bark: Review of approaches from extractives to materials for higher-added value products. *Materials Today Sustainability*. v.29, 101074, 2025. <https://doi.org/10.1016/j.mtsust.2025.101074>.

TSUCHIKAWA, S.; INAGAKI, T.; MA, T. Application of near-infrared spectroscopy to forest and wood products. *Current Forestry Reports*. v.9, p.401-412, 2023. <https://doi.org/10.1007/s40725-023-00203-3>

ZHANG, X.; YU, H.; LI, B.; LI, W.J.; LI, W.; BAO, C. 2014. Discrimination of *Pinus yunnanensis*, *P. kesiya* and *P. densata* by FT-NIR. *Journal of Chemical and Pharmaceutical Resources*. v.6, n.4, p.142-149, 2014.

ZHOU, Z.; RAHIMI, S.; AVRAMIDIS, S. On-line species identification of green hem-fir timber mix based on near infrared spectroscopy and chemometrics. *European Journal of Wood and Wood Products*. v.78, p.151-160, 2020. <https://doi.org/10.1007/s00107-019-01479-8>