Article

Machine Learning Algorithms and Nondestructive Methods for Estimating Wood Density in Planted Forest Trees

Rafael Gustavo Mansini Lorensani [*] and Raquel Gonçalves

Article

Machine Learning Algorithms and Nondestructive Methods for Estimating Wood Density in Planted Forest Trees

Rafael Gustavo Mansini Lorensani [1,*] and Raquel Gonçalves [2]

[1] Valora Madeira Classificação e Inspeção Ltda, São Paulo, Brazil

[2] Laboratory of Nondestructive Testing, College of Agriculture Engineering (FEAGRI), State University of Campinas (UNICAMP), São Paulo, Brazil

* Correspondence: rafael@valoramadeira.com

Abstract: Inferring forest properties is crucial for the timber industry, enabling efficient monitoring, predictive analysis, and optimized management. Nondestructive testing (NDT) methods have proven to be valuable tools for achieving these goals. Recent advancements in data analysis, driven by machine learning (ML) algorithms, have revolutionized this field. This study analyzed 492 eucalyptus trees, aged 3 to 7 years, planted in São Paulo, Brazil. Data from forest inventories were combined with results from ultrasound, drilling resistance, sclerometric impact, and penetration resistance tests. Seven machine learning algorithms were evaluated to compare their generalization capabilities with conventional statistical methods for predicting basic wood density. Among the models, Extreme Gradient Boosting (XGBoost) achieved the highest accuracy, with a coefficient of determination (R^2) of 89% and a root mean square error (RMSE) of 10.6 kg·m⁻³. In contrast, the conventional statistical model, using the same parameters, yielded an R^2 of 33% and an RMSE of 26.4 kg·m⁻³. These findings highlight the superior performance of machine learning in nondestructive inference of wood properties, paving the way for its broader application in forest management and the timber industry.

Keywords: basic density; machine learning; ultrasound; drilling resistance

1. Introduction

Despite advances in genetic engineering and cloning, the properties of wood are still influenced by various factors (climate, soil, altitude, etc.) related to the development of trees. These influences result in significant variability in the wood produced in the forest, making it difficult to establish important reference values for the production line of forest companies linked to pulp and paper, wood-based products (fiberboards and particleboards), as well as roundwood or processed wood used in solid or engineered forms (glued laminated timber or cross-laminated timber) in construction.

The prediction of tree properties in the forest has been a recurring topic in both national and international publications, with non-destructive techniques being a viable alternative to this challenge. By not affecting the material, these techniques allow tests to be repeated during the tree's growth. They can be performed directly on standing trees [1,2], and since no sampling or tree cutting is required, they are more economical, simpler, and faster, enabling increased sampling. Thus, several research groups have been developing studies using various nondestructive techniques aimed at inferring properties such as wood density (basic or apparent) and stiffness [3–12]. These nondestructive methods include those that use wave propagation (ultrasound and stress waves), resistance to drilling, resistance to penetration, mass spectrometry, among others. Regardless of the type of equipment or technique used, the focus of the research is to correlate the measured parameters in trees with wood properties, with statistical modeling being the tool typically used to obtain the

relationship between variables to predict results. Classical statistical tools are no longer the most suitable when the volume of data is very large, and machine learning algorithms are currently recommended for such cases.

The adoption of non-destructive techniques as a tool for evaluating the quality of wood produced in forests stems from the possibility of obtaining correlations between the responses of these methods and the physical and mechanical properties of the fibers, which are used for the selection and classification of forest products. Thus, the adoption of such techniques in the wood production chain represents a fast and effective means of assessing wood quality, with the potential to enhance inference models already employed by companies in the forestry sector [13–15].

The literature review indicated that in Brazil, studies related to machine learning are more strongly focused on estimating tree volume [16–20]. In other countries, [21] and [22] researches were the closest to our objective, as the authors studied different machine learning algorithms in inferring density, with superior results to those obtained using classical multiple regression. In field trials, [21] and [22] used dendrometric measurements of trees (height and diameter at breast height (DBH), NDT techniques (stress wave propagation and resistance to drilling (only [21]), and extracted cores with an increment borer for density measurement using X-ray densitometry.

The main objective of this research was to evaluate whether machine learning algorithms, utilizing data collected from standing trees (forest inventory and non-destructive inspections), can provide accurate predictions of wood basic density compared to conventional statistics methods.

2. Materials and Methods

The sampling comprised 491 eucalyptus trees planted in the state of São Paulo, Brazil, aged between 3 and 7 years. In addition to the data collected in the conventional inventory of the partnering company (Suzano S.A.) (diameter at breast height, total height, and pylodin penetration), data from ultrasonic wave propagation tests in the longitudinal and radial directions, drilling resistance, sclerometric impact, and sample retrieval with the Pressler probe were added (Figure 1). From each plot, 3 trees were chosen according to their DBH class (1: 8±4cm, 2: 12±4cm, 3: 16±4cm) and cut for determination of basic density in the laboratory of the partnering company. The data were compiled into a single database, containing the inventory results and additional tests, to which 7 machine learning algorithms were applied, selected with the aim of predicting basic density from the other variables, following a sequence of steps (Figure 2). The basic density of the trees that were not cut was estimated using a model adopted by the partnering company.

(**a**)   (**b**)

Figure 1. Ultrasonic wave propagation test in the longitudinal direction (a), drilling resistance test (b), sclerometric impact test (c), and sample retrieval by Pressler probe (d).

Figure 2. Sequence of steps performed during the modeling of machine learning algorithms.

The process of splitting the data into training and testing sets was carried out to ensure that the models were trained with distinct datasets in each training round. This approach increases their generalization capacity and reduces the likelihood of overfitting. The normalization step was critical to prevent dimensional differences between variables from being incorporated into the models. Moreover, some algorithms operate more effectively with normalized data, enhancing their performance.

The feature selection process aimed to identify the most relevant variables for predicting basic density in eucalyptus forests, optimizing model performance and reducing data dimensionality. Initially, the ExtraTreeRegressor algorithm was used to calculate feature importance based on the variance reduction provided by tree splits. Subsequently, other methods, such as Random Forest and XGBoost, were applied to validate the results and ensure robustness in the selection process. The most relevant variables were then selected to form the final set of features used in training the predictive models.

To prevent data leakage and ensure model reliability, the feature importance analysis was conducted exclusively on the training set after splitting the data. Only the variables selected based on the training data were used in the testing set. This approach ensures that model evaluation reflects performance on truly unseen data, simulating practical application scenarios [23].

All data mining methods were configured following the guidelines of [24–26] for hyperparameter tuning. Individual adjustments were made for each model to mitigate the risks of overfitting and underfitting. During hyperparameter optimization, different seeds were used for initializing evaluations, allowing the assessment of varying model behaviors. A 10-fold cross-validation strategy (cv=10) was also employed. Overfitting occurs when a model fits the training data exceptionally well but fails to generalize to new data. Conversely, underfitting happens when the

model cannot capture relationships between the variables in the training set, causing the process to terminate prematurely, even before testing [26].

The variables selected by machine learning algorithms were also evaluated using conventional statistical software (Statgraphics Centurion XV 15.1.02) to identify the best model for inferring wood basic density. This process involved applying classical statistical modeling techniques, such as multiple regression and models automatically suggested by the software, aiming to maximize the coefficient of determination (R2) and minimize the root mean square error (RMSE).

After the prediction step, each model was evaluated using the following metrics: coefficient of determination (R^2), adjusted coefficient of determination (adjusted R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

3. Results

3.1. Laboratory and Field Results

The mean values, standard deviation, coefficient of variation, skewness, and kurtosis of the numerical variables evaluated in this research are presented in Table 1. As mentioned previously, we can notice the difference between the scales of the variables and that is why it is important to carry out the normalization step.

Table 1. Mean values, standard deviation, coefficient of variation, skewness, and kurtosis of the parameters.

Parameters	Mean	Stand. Dev.	C.V.	Skewness	Kurtosis
Age [years]	5,31	1,48	27,49%	-3,4048	-5,4858
DBH [cm]	13,86	3,40	24,53%	1,031	0,457
H [m]	21,71	5,18	23,86%	-1,048	-3,688
VL [m.s^{-1}]	4317,61	566,70	13,13%	-4,367	2,101
VR [m.s^{-1}]	1946,18	278,91	14,33%	1,849	10,786
SI [%]	26,51	5,95	22,44%	5,777	31,211
DR [%]	24,30	3,35	13,79%	-1,311	-0,498
BdSU [kg.m^{-3}]	441,90	32,24	7,30%	3,622	-3,341
BdV [kg.m^{-3}]	386,34	34,12	8,83%	3,063	-0,407
VITCC [m^3]	0,165	0,108	65,45%	9,664	4,362
VITCCR [m^3]	231,24	93,32	40,36%	-0,961	-5,416
IMATCC [m^3/ha.year]	46,28	8,66	18,71%	-3,551	-1,945

Note: DBH = diameter at breast height, H = height, VL = longitudinal ultrasonic pulse velocity, VR = radial ultrasonic pulse velocity, SI = sclerometric impact, DR = drilling resistance, BdSU = basic density provided by Suzano, BdV = basic density obtained by borer, VITCC = total volume with bark, VTCCR = total volume of the production unit, IMATCC = annual mean increment.

3.2. Feature Selection

During the feature selection stage, the following variables were selected: DBH, H, VL, DR, AGE, and IMATCC. AGE, DBH, H, and IMATCC are already part of the inference models employed by companies. Conversely, VL and DR stood out as the non-destructive methods with the highest levels of importance within the models. It is also well-known that the individual use of DBH and H in models can lead to autocorrelation issues, as observed in Table 2. Variance Inflation Factors (VIF) were calculated for each variable, and although there is no definitive consensus on acceptable VIF limits, we adopted the criterion that values exceeding 5 should be excluded or transformed. To address multicollinearity, these variables were combined (H/DBH).

Table 2. Variance Inflation Factors of the variables, before and after the creation of a new variable.

Original variables	Original VIF	Combined variables	Combined VIF
Const.	133,22	Const.	132,02

Age	3,09	Age	1,84
DBH	6,43	H/DBH	1,87
H	13,55	-	-
VL	1,60	VL	1,56
DR	1,89	DR	1,39
IMATCC	1,80	IMATCC	1,39

3.3. Hyperparameter Optimization

The hyperparameter tuning process was conducted using multiple random seeds to initialize the evaluations and a 10-fold cross-validation strategy (cv=10) (Table 3). The coefficient of variation (CV) for the evaluation metrics – mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), and adjusted coefficient of determination (R^2 adj) – reached up to 21% across different random seeds (Table 4).

Table 3. Optimized hyperparameters for each algorithm evaluated.

Model	Optimized hyperparameters
KNN	n_neighbors: 3.0
Decision Tree	max_depth: 11.0, min_samples_split: 3.0, min_samples_leaf: 6.0
Random Forest	max_depth: 11.0, min_samples_split: 6.0, min_samples_leaf: 2.0, n_estimators: 225.0
Gradient Boosting	max_depth: 8, min_samples_split: 8, min_samples_leaf: 6, n_estimators: 175, learning_rate: 0.1, max_features: None, subsample: 1.0
Extreme Gradient Boosting	max_depth: 3.0, n_estimators: 100.0, learning_rate: 0.2, subsample: 0.8, colsample_bytree: 0.8, min_child_weight: 1.0, gamma: 5.0
SVM	kernel: rbf, C: 100, gamma: 1, epsilon: 0.5
ANN	hidden layer sizes: (200,), max_iter: 600, alpha: 0.0001, batch_size: 32

Table 4. Mean and coefficient of variation of the evaluation metrics for the algorithms.

Algorithm	RSME	MAE	R2	R2 adj.
KNN	20,77 (7,0%)	14,87 (8,0%)	56% (11,1%)	54% (11,8%)
Decision Tree	14,96 (14,8%)	8,63 (14,9%)	77% (8,2%)	76% (8,6%)
Random Forest	12,91 (9,6%)	8,53 (10,2%)	83% (4,4%)	82% (4,6%)
Gradient Boosting	12,18 (12,7%)	7,84 (10,3%)	85% (4,6%)	84% (4,8%)
Xtreme Gradient Boosting	12,05 (9,0%)	8,07 (7,3%)	85% (3,1%)	85% (3,2%)
Support Vector Machine	20,79 (5,5%)	14,85 (7,5%)	56% (8,8%)	54% (9,4%)
Artificial Neural Network	26,19 (2,2%)	21,58 (2,75%)	28% (18,9%)	25% (21,5%)

3.4. Classical Statistics

The model obtained by classical statistics (Equation 1) was evaluated by parametrical statistics (Statgraphics Centurion XV 15.1.02), using the same variables previously selected. In this study, the classical statistical model was selected using the best-fit model proposed by the Statgraphics software, ensuring the most appropriate approach was utilized for comparison.

$$DbSU = 300,823 + 6,11588*AGE + 4,89593*H/DAP + 0,0170558*VL + 0,0270439*DR - 0,832398*IMATCC \qquad (1)$$

3.5. Optimal Metrics

Table 5 presents the best values obtained by the algorithms evaluated in this research with respect to the assessment metrics.

Table 5. Evaluation metrics of the best assessed algorithms.

Algorithm	RMSE	MAE	R2	R2 adj
Classical Statistics	26,40	21,66	33%	32%
KNN (K-Nearest Neighbor)	18,82	12,85	66%	65%
Decision Tree	12,38	7,34	81%	80%
Random Forest	11,59	8,14	83%	82%
Gradient Boosting	10,20	7,10	87%	86%
Extreme Gradient Boosting	10,64	7,67	89%	88%
SVM (Superior Vector Machine)	18,68	13,25	67%	66%
ANN (Artificial Neural Network)	25,53	20,62	19%	16%

In addition to the metrics (Table 5), two graphs were generated for each model: one showing the dispersion of predicted values versus observed values, and the other showing the dispersion of residuals (Figures 3 and 4). The points representing the estimated and actual values are concentrated around the 45° line, indicating their proximity (Figure 3).
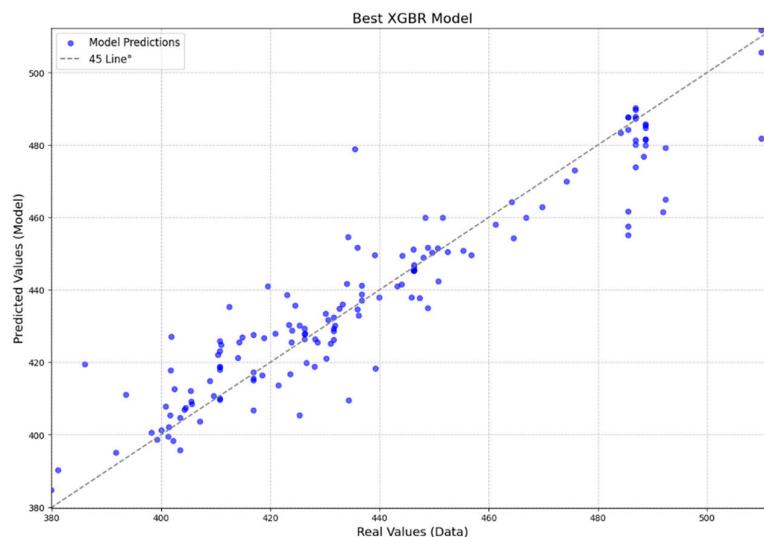
Figure 3. Results predicted vs observed graph for the Extreme Gradient Boosting algorithm.
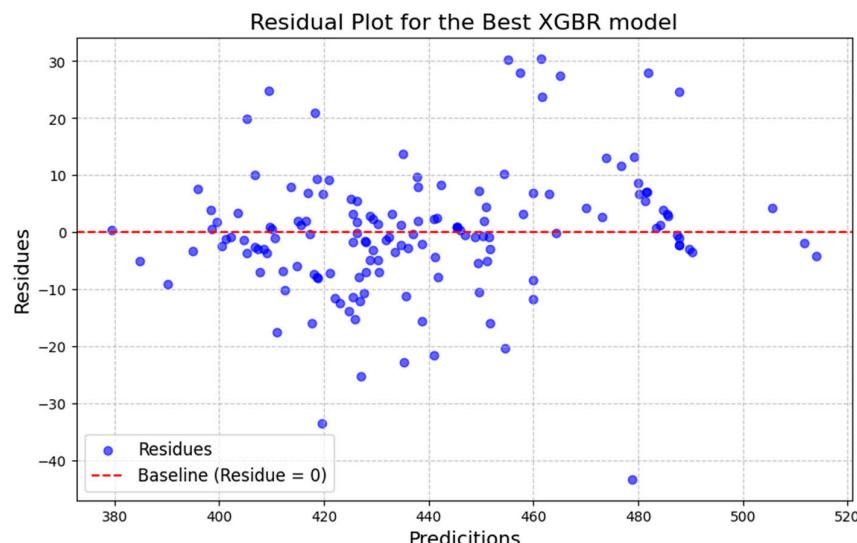

Figure 4. Residuals graph for the Extreme Gradient Boosting algorithm.

4. Discussion

4.1. Laboratory and Field Results

Some parameters do not have a normal distribution because skewness and/or kurtosis values are outside the expected range of -2 to 2 for this type of distribution (Table 1). For evaluations involving conventional parametric statistics, it sometimes necessary to attempt transformations, as for example in the case of statistics related to mean comparisons. In the case of multiple linear regressions, normality must be present in the residuals (errors) of the model, not necessarily in the independent or dependent variables [27,28].

4.2. Hyperparameter Optimization

During the hyperparameter optimization stage, we observed that the maximum coefficient of variation was 21% among all evaluated seeds. This result indicates moderate variability in the model's performance, suggesting that random initialization and variations in data splits may have a noticeable, though not critical, impact. Nonetheless, the CV values are within an acceptable range, indicating that the models provide reasonably consistent outcomes. This variability emphasizes the importance of using multiple seeds and cross-validation to ensure reliable and generalizable results. Overall, the models remain suitable for the proposed task of basic density prediction in eucalyptus forests [26].

4.3. Classical Statistics

The use of predictive models evaluated through classical statistical techniques is a well-established method. Thus, to confirm the hypothesis of this research, we obtained Equation 1. This approach aligns with methodologies employed in other studies. For instance, [29] evaluated Eucalyptus camaldulensis samples using ultrasound and developed density inference models with a determination coefficient of 34%. Similarly, [30] assessed three conifer species using drilling resistance, which is also one of the methods evaluated in this research, achieving basic density inference models with determination coefficients ranging from 25% to 52%.

4.4. Optimal Metrics

Comparing the MAE to the mean laboratory value of basic density contextualizes the error, highlighting its practical relevance and confirming the model's robustness in handling the natural variability of basic density. Evaluating the MAE as a percentage of the mean basic density obtained from laboratory tests, values ranged from 1.61% to 4.90% (Table 5). These results indicate low estimation errors relative to the coefficient of variation of basic density. Such findings suggest that the model effectively captures the variability of the data and provides accurate predictions that align with the observed dispersion of actual values.

When modeling a variable in relation to others, it is essential to always consider the composition of the dataset, its size, and how the variables within the dataset relate to each other. It is important to highlight that this study utilized a dataset consisting of trees planted at varying spacings and locations—factors that significantly influence density. However, these variables were not explicitly modeled in this phase of the research. While the model could potentially be improved by adding these factors to the dataset, this was not the objective of this paper, which aims to compare conventional techniques and machine learning approaches in modeling data for density inference.

When graphically analyzing the dispersion of predicted versus actual data (Figure 3), where the intersection of the data is concentrated near the 45-degree line dividing the graph, it suggests that the model is capable of generalizing the data effectively and predicting new data with efficiency. Furthermore, the residuals are randomly distributed and do not show a trend (Figure 4). This indicates that the model does not exhibit bias, overfitting, or underfitting issues, further supporting its robustness and reliability [25].

[20] presented predicted vs. observed plots and residual analyses that were comparable to those obtained in this study (Figures 3 and 4). Although their study utilized a database approximately 4.5 times larger than ours—resulting in a more robust training dataset—and included additional variables such as basal area and number of stems per hectare, it did not incorporate nondestructive testing results. Despite working with a smaller database, our study achieved metrics of a similar magnitude, demonstrating the effectiveness of the selected variables and modeling approaches.

For species different than those examined in this research (e.g., Douglas-fir, fir hybrids, acacias, and species from the Brazilian Cerrado), with tree ages ranging from 32 to 50 years, the literature reports R^2 values between 50% and 80%. These studies employed various algorithms and architectures such as Artificial Neural Networks (ANN), Classification and Regression Trees (CART), and Support Vector Machines (SVM) [21,22,31,32]. These performance ranges are closer to those obtained in the present study, where R^2 values ranged from 19% to 89% (Tables 2 and 3).

The results from the literature make it clear that the database size and variable selection significantly impact model performance, and that the incorporation of nondestructive testing methods provides a novel contribution of this study to the field. In addition to enhancing predictive accuracy, nondestructive testing data introduce information about the quality of the fiber produced, which is not captured by traditional models based solely on mass and volumetric variables [13; 14; 15]. This added dimension positions nondestructive methods as valuable tools for advancing wood quality assessments, addressing gaps in current modeling approaches.

5. Conclusions

Machine learning algorithms, when applied to databases derived from measurements on standing trees and enhanced with data from nondestructive techniques, enable the development of suitable models for accurately inferring the basic density of wood.

This study highlighted the superior performance of machine learning models compared to those derived from classical statistical approaches. While classical methods offer simplicity and interpretability, machine learning models demonstrated significantly higher predictive accuracy and better generalization. Moreover, integrating data from nondestructive methods into the models not only improved performance but also incorporated valuable information about fiber quality—an aspect overlooked by traditional mass and volumetric models.

These findings underscore the importance of leveraging advanced modeling techniques in conjunction with robust datasets and interpretable frameworks. Such an approach enhances confidence in machine learning applications and supports their integration into forest management and industrial decision-making, offering a pathway toward more efficient and sustainable use of timber resources.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gonçalves, R. et al. Evolution of acoustical, geometrical, physical, and mechanical parameters from seedling to cutting age in Eucalyptus clones used in the pulp and paper industries in Brazil. Forest Products Journal, v. 69, n. 1, p. 5–16, 2019. DOI: 10.13073/FPJ-D-17-00013.

2. Gonçalves, R. et al. Modeling of wood properties from parameters obtained in nursery seedlings. Canadian Journal of Forest Research, v. 48, n. 6, p. 621–628, 2018. DOI: 10.1139/cjfr-2017-0393.

3. Gendvilas, V. et al. Friction correction when predicting wood basic density using drilling resistance. Holzforschung, v. 75, n. 6, p. 508–516, 1 jun. 2021. DOI: 10.1515/hf-2020-0156.

4. Nickolas, H. et al. Application of resistance drilling to genetic studies of growth, wood basic density and bark thickness in Eucalyptus globulus. Australian Forestry, v. 83, n. 3, p. 172–179, 2 jul. 2020. DOI: 10.1080/00049158.2020.1808276.

5. De Pádua, F. A. et al. Recommendation of non-destructive sampling method for density estimation of the Eucalyptus wood. Maderas: Ciencia y Tecnologia, v. 21, n. 4, 2019. DOI: 10.4067/S0718-221X2019005000412.

6. Kharrat, W. et al. Intra-ringwood density and dynamic modulus of elasticity profiles for black spruce and jack pine from X-ray densitometry and ultrasonicwave velocity measurement. Forests, v. 10, n. 7, 1 jul. 2019. DOI: 10.3390/f10070569.

7. Liu, F. et al. Use of time-of-flight ultrasound to measure wave speed in poplar seedlings. Forests, v.10, n. 8, 2019. DOI: 10.3390/f10080682

8. Downes, G. M. et al. Application of the IML Resistograph to the infield assessment of basic density in plantation eucalypts. Australian Forestry, v. 81, n. 3, p. 177–185, 3 jul. 2018. DOI: 10.1080/00049158.2018.1500676.

9. Fundova, I.; Funda, T.; Wu, H. X. Non-destructive wood density assessment of Scots pine (Pinus sylvestris L.) using Resistograph and Pilodyn. PLoS ONE, v. 13, n. 9, 1 set. 2018. DOI: 10.1371/journal.pone.0204518.

10. Carrillo, I.; Valenzuela, S.; Elissetche, J. P. Comparative evaluation of Eucalyptus globulus and E. nitens wood and fibre quality. IAWA Journal, v. 38, n. 1, p. 105–116, 2017. DOI: 10.1163/22941932-20170160.

11. Soriano, J.; Da Veiga, N. S.; Martins, I. Z. Wood density estimation using the sclerometric method. European Journal of Wood and Wood Products, v. 73, n. 6, p. 753–758, 1 nov. 2015. doi.org/10.1007/s00107-015-0948-3.

12. Sandoz, J. L. Standing Tree Quality Assessments Using Ultrasound. Acta Horticulturae, n. 496, p. 269–278, set. 1999. doi.org/10.1016/j.compag.2023.107816.

13. Proto, A. .; Macrì, G.; Bernardini, V.; Russo, D.; Zimbalatti, G. 2017. Acoustic evaluation of wood quality with a non-destructive method in standing trees: a first survey in Italy. iForest 10: 700-706. – doi: 10.3832/ifor2065-010

14. Ondrejka, V.; Gergeľ, T.; Bucha, T.; Pástor, M. 2020. Innovative methods of non-destructive evaluation of log quality. Central European Forestry Journal. 66. 1-11. 10.2478/forj-2020-0021.

15. Jones, G.; Liziniewicz, M.; Lindeberg, J.; Adamopoulos, S. Non-Destructive Evaluation of Downy and Silver Birch Wood Quality and Stem Features from a Progeny Trial in Southern Sweden. Forests 2023, 14, 2031. https://doi.org/10.3390/f14102031

16. Aragão, M. DE A.; Santos, J. S.; Da Silva, M. L. M. Técnica de mineração de dados aplicada a estimação de volume de árvores de eucalyptus. 5a Semana de Engenharia Florestal da Bahia. Anais...Vitória da Conquista: mar. 2018.

17. Lima, E. de S. et al. Classification of the initial development of eucalyptus using data mining techniques. Cerne, v. 23, n. 2, p. 201–208, 2017. doi.org/10.1590/01047760201723022296.

18. Cordeiro, M. A. et al. Estimativa do volume de Acacia mangium utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. Pesquisa Florestal Brasileira, v. 35, n. 83, p. 255, 30 set. 2015. DOI: 10.4336/2015.pfb.35.83.596.

19. Binoti, M. L. M. DA S. et al. Redes neurais artificiais para estimação do volume de árvores. Revista Árvore, v. 38, n. 2, p. 283–288, abr. 2014. DOI: 10.1590/S0100-67622009000600016.

20. Leite, H. G. et al. Redes Neurais artificiais para a estimação da densidade básica da madeira. Scientia Forestalis/Forest Sciences, v. 44, n. 109, p. 149–154, 1 mar. 2016. DOI: 10.18671/scifor.v44n109.14.

21. Demertzis, K. et al. Machine learning use in predicting interior spruce wood density utilizing progeny test information. Neural Computing and Applications, v. 28, n. 3, p. 505–519, 1 mar. 2017. doi.org/10.1007/s00521-015-2075-9.

22. Iliadis, L. et al. Predicting Douglas-fir wood density by artificial neural networks (ANN) based on progeny testing information. Holzforschung, v. 67, n. 7, p. 771–777, 1 out. 2013. DOI: 10.1515/hf-2012-0132.

23. Nassar, O. 2023. Data Leakage in Machine Learning. 10.13140/RG.2.2.27468.59528.

24. Plas, J. V. Python Data Science Handbook. 2016. ISBN: 9781491912058. Disponível em: <https://jakevdp.github.io/PythonDataScienceHandbook/>.

25. McKinney, W. 2012. Python for data analysis (3rd. ed.). ISBN: 978-1-098-10403-0. O'Reilly Media, Inc.

26. Goldschimidt, R.; Passos, E. Data mining: um guia prático. Rio de Janeiro: Campus, 2005. DOI: 10.21529/RESI.2006.0501011.

27. Sainani, K.L. 2012, Dealing With Non-normal Data. PM&R, 4: 1001-1005.https://doi.org/10.1016/j.pmrj.2012.10.013

28. Chowdhury, S.; Lin, Y.; Liaw, B.; & Kerby, L. 2021. Evaluation of Tree Based Regression over Multiple Linear Regression for Non-normally Distributed Data in Battery Performance. 2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA), 17-25.

29. Melo, R. R.; Barbosa, K. T.; Beltrame, R.; Acosta, A. P.; Pimenta, A. S.; Mascarenhas, A. R. P. 2020. Ultrasound to determine physical-mechanical properties of Eucalyptus camaldulensis wood, Wood Material Science & Engineering, DOI: 10.1080/17480272.2020.1830435.

30. Yao, J.; Zhao, Y.; Lu, J.; Liu, H.; Wu, Z.; Song, X.; Li, Z. Research on the Wood Density Measurement in Standing Trees though the Micro Drilling Resistance Method. Forests 2024, 15, 175, DOI: 10.20944/preprints202312.1940.v1.

31. Iglesias, C. et al. Influence of heartwood on wood density and pulp properties explained by machine learning techniques. Forests, v. 8, n. 1, 1 jan. 2017. DOI: 10.3390/f8010020.

32. Silva, J. P. M.; Cabacinha, C. D.; Assis, A. L.; Monteiro, T. C.; Araújo Júnior, C. A.; Maia, R. D. 2018. Redes neurais artificiais para estimar a densidade básica de madeiras do cerrado. Pesquisa Florestal Brasileira. 38. 10.4336/2018.pfb.38e201801656.