Engineering Structures 254 (2022) 113826

Contents lists available at ScienceDirect

Engineering Structures

journal homepage: www.elsevier.com/locate/engstruct

Assessment of machine learning algorithm-based grading of *Populus* x *euramericana* I-214 structural sawn timber

Luis Acuña-Rello [a], Eleana Spavento [b,*], Milagros Casado-Sanz [a], Luis-Alfonso Basterra [a], Gamaliel López-Rodríguez [a], Gemma Ramón-Cueto [a], Enrique Relea-Gangas [a], Leandro Morillas-Romero [a], David Escolano-Margarit [a], Roberto D. Martínez [a], José Antonio Balmori [a]

[a] *Timber Structures and Wood Technology Research Group, University of Valladolid, C/Plaza de Santa Cruz, 8, 47002 Valladolid, Spain*

[b] *Wood Research Laboratory (LIMAD, as per initials in Spanish), School of Agrarian and Forestry Sciences, National University of La Plata, Av. 60 S/N, 1900 La Plata, Argentina*

ARTICLE INFO

Keywords:
Poplar
Timber grading
Defects
Sawing systems
Non-destructive testing
Strength class

ABSTRACT

The efficiency of visual grading standards applied to structural timber is often inappropriate, and timber properties are either under or over-graded. Although not included in the current UNE 56544 visual grading standard, machine learning algorithms represent a promising alternative to grade structural timber. The general aim of this research was to compare the performance of machine learning algorithms based on visual defects, non-destructive techniques and sawing systems ("cut type") with UNE 56544:1997 visual grading in order to predict the qualifying efficiency of *Populus* x *euramericana* I-214 structural timber. Visual evaluation, ultrasound and vibrational non-destructive testing, and sawing systems register (radial, tangential and mixed) were applied to characterize 945 beams. In addition, in order to retrieve actual physical-mechanical values, density and static bending destructive testing (EN-408:2011 + A1:2012) was also carried out. Several machine learning algorithms were then used to grade the beams, and their predictive accuracy was compared with that of visual grading. To do so, three scenarios were considered: a first scenario in which only visual variables were used; a second scenario in which "cut type" variables were also included; and a third scenario in which additional non-destructive variables were considered. Results showed a poor level of performance of UNE 56544:1997, with an apparent mismatch between the strength values assigned for each visual grade (established by the EN 338 standard) and the actual values. On the opposite, all algorithms performed better than visual grading and may thus be deemed as promising timber strength grading tools.

1. Introduction

Timber physical and mechanical properties, growth conditions and defects need to be assessed to ensure its efficient structural use and safety. In this sense, visual grading is the oldest and most widely used method for assignment strength grades to sawn timber. Visual grading involves inspecting, recording and evaluating defects, whose presence and magnitude define the structural visual grade of the pieces. It is carried out following grading rules usually defined in national standards, which allow safe and economic grading results. In general, the national visual grading standards tend to be optimized for the needs of the publishing country and take into account the species, provenance, growth conditions, cross-sections and silvicultural treatments with different efficiency and different assignments to structural strength classes [1–3].

The first standard for visual grading of structural sawn timber in Spain was UNE 56544:1997 [4]. It was applied to softwood and hardwood species grown in Spain (viz., *Pinus sylvestris, Pinus pinaster*, *Pinus radiata*, *Eucalyptus globulus* and *Populus* spp.). However, the current UNE 56544 version (2011) is only applicable to conifer species [5]. Therefore, this study will use the 1997 version of UNE 56544, given that it is the only standard that includes *Populus* spp. in its scope.

The UNE 56544:1997 standard [4] defines two grades for structural timber (ME-1 and ME-2), taking into account the presence and size of

* Corresponding author.
E-mail address: eleana.spavento@agro.unlp.edu.ar (E. Spavento).

https://doi.org/10.1016/j.engstruct.2021.113826
Received 2 June 2021; Received in revised form 12 November 2021; Accepted 29 December 2021
Available online 20 January 2022

defects such as face and edge knots, checks and splits, pitch pockets, juvenile and reaction wood, slope of grain, wane, pith and biological alterations (decay, blue stain, moulds, stains and hole insects). Nonetheless, wood is an anisotropic material, and, therefore, it has a differential behaviour as a function of sawing systems (radial, tangential or mixed). Wood shrinks (swells) are primarily present in the direction of the annual growth rings (tangentially), about half as much across the rings (radially), and only slightly present along the grain (longitudinally). The combined effects of radial and tangential shrinkage can distort the shape of wood pieces because the difference in shrinkage and the curvature of annual rings result in warps (bow, crook, cup, twist). This is especially important during the drying process because it can influence structural timber use [6–8]. Even though sawing systems and warps are not included in the visual grading specifications of ME-1 and ME-2 in this version of the standard, given the influence of wood anisotropy, in this work, the sawing systems, ring width and warps will be taken into account.

Furthermore, UNE 56544:1997 [4] assigns, by each visual grade and species, structural strength classes (C class, given in EN 338 standard [9]), based on the most important physical and mechanical (static bending destructive test) values obtained: density, modulus of elasticity-MOE (stiffness), and modulus of rupture-MOR (bending strength). However, to account for the variation of these values, the classification is based on the so-called characteristic values, which are the fifth percentile value for density and MOR, and mean values for MOE [10]. In the case of *Populus* spp. timber, UNE 56544:1997 standard correlates visual grade ME-1 with strength class C18 and visual grade ME-2 with strength class C14.

To facilitate the exchange of structural timber between different markets and homogenize the national visual grades in all European countries, EN 1912 standard [11] lists how national visual grades are related to strength classes defined in EN 338 [2,9,12]. Unfortunately, *Populus* spp. from Spain has never been included in any EN 1912 standard versions because the number of essays did not suffice to contrast them at the European level. In this regard, to include timber belonging to a species, provenance and grade into EN 1912 standard, extensive experience of use or essay results should be justified under EN 384, from which the characteristics values are obtained in order to assign a strength class according to EN 338 [13].

Even though visual grading is the most widely used method for assigning structural strength classes to sawn timber, the qualifying efficiency of different visual grading standards applied to structural timber is often inappropriate, and timber properties are either under or over-graded. Non-destructive techniques (NDTs) are adequate to assist visual grading, being a reliable and straightforward method to evaluate the performance and estimate physical and mechanical characteristics of samples of wood or even wood structures in service [14–16]. These NDTs have been developed and used over the last decades [17]. A literature review about the use of techniques on Spanish timber can be found on [12], which presents a panorama of works about density, MOR and MOE estimation from acoustic (ultrasound and stress wave), vibrations and probing techniques. Some authors have also put forward the use of mixed techniques for structural grading and mechanical properties estimation based on a combination of NDTs and visual parameters [17,18].

On the other hand, machine learning (ML) uses artificial intelligence algorithms that, using computers, improve the performance of tasks based on measured data. ML can be used to make predictions about future data and make decisions that are rational given these predictions [19]. There are many ML algorithms, among which Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Decision Tree, Naive Bayes, Artificial Neural Network (ANN or nnet) are some of the most popular. Some of these algorithms have been used in studies on timber in the literature. For instance, SVM was used for knots detection on different tropical timbers, with promising results to improve the classifier [20]. SVM was also used to identify small samples of timber species in combination with infrared spectroscopy (IR), finding that this technology led to better predicting results than Cluster Analysis and Bayes Discriminant [21]. Likewise, Dos Santos et al. 2021 [22], using a combination of near-infrared spectroscopy (NIR) and machine learning techniques (SVM, KNN and partial least squares discriminant analysis), were able to recognize wood from the “Louros” group from the Brazilian Amazon.

Notwithstanding the above results, neural networks are the most widely used ML algorithms in connection with sawn timber. They have been used to grade wood defects (knots) [23], to predict MOR and MOE [24] and compression strength [25] on heat-treated wood, to predict the bonding strength of the wood joints pressed under different conditions [26], and to estimate wood resistance [27], among other applications, with consistently good prediction and classification performances. In Spain, neural networks were used to predict Spanish timber’s physical and mechanical properties in combination with NDTs and visual grading [12]. Another study compared the performance of neural networks for the visual grading of Spanish *Pinus nigra* and *Pinus sylvestris* timber according to UNE 56544/1M:2003, concluding that UNE 56544 is overly conservative for those two species, underrating their physical-mechanical properties, and that neural networks hold great potential as timber grading tools [28].

However, to the best of the authors’ knowledge, ML algorithms have not been assayed to grade *Populus* spp. Spanish structural timber. Hence, taking into account that *Populus* spp. represent a large surface of Spanish forested area (over 120,000 ha, according to the Land Use and Crop Yield Survey [29]), and that their timber is not included in current UNE 56544 and EN 1912 standards, ML may be regarded as an exciting strategy to grade such structural timber. The traditional use of this timber is associated with low value-added applications [30–35]. However, there is room for poplar valorization, particularly in the construction and building sector, through the development of engineering products with higher added value (studies on this topic with promising results were carried out by [36,37]).

In this context, it would be interesting to add ML methods to current international standards and compare and evaluate the influence of defects on the physical and mechanical properties of *Populus* x *euramericana* (Dode) Guinier I-214 for structural purposes. The importance of this clone lies in that it is widely spread in the world [38], representing over 50% of the total amount of poplar available in Spain. This has attracted interest in studying its physical-mechanical properties for structural purposes to fulfil requirements established by the strength classes standard EN-338. Also, there is an increasing interest in the *Populus* x *euramericana* I-214 timber incorporation into the current European grading standard in force (EN-1912). To attain this goal, it is essential to evaluate the behaviour of this timber by UNE-56544 standard and by visual grading rules in combination with other NDT techniques and with sawing variables to assess its performance.

The aims of this paper were:

- to evaluate the qualifying efficiency of the Spanish visual grading standard (UNE-56544:1997) applied to *Populus* x *euramericana* (Dode) Guinier I-214 structural timber.
- to evaluate, through ML algorithms, the qualifying efficiency of visual variables (defects, UNE 56544:1997) along with two NDTs (ultrasound and longitudinal vibration analysis) and three sawing systems (radial, tangential or mixed) for this structural timber.
- to compare the performance of ML algorithms with that of the visual grading standard currently in use in terms of their ability to predict the qualifying efficiency of this structural timber.

2. Materials and methods

2.1. Structural timber

Nine hundred forty-five timber beams of *Populus* x *euramericana* I-214 were selected from 20-year old trees from five areas of Castilla-y-León (NW Spain), with the following nominal section sizes (Table 1).

2.2. Visual grading

The specimens were numbered, stacked and conditioned in the laboratory to achieve an equilibrium moisture content of 12 ± 2% (determined by a digital hygrometer, model Testo 606-1). Their defects were measured following the European standard EN-1310:1997 [39], attending to three categories: timber anatomy (Fig. 1), sawing (Fig. 2), and drying process, i.e., warps (Fig. 3).

Table 1
Dimensions of the *Populus* x *euramericana* I-214 studied timber beams.

	Sample A	Sample B	Sample C	Sample D	Sample E
Thickness (mm)	50	52	80	80	100
Width (mm)	150	135	120	150	200
Length (mm)	3050	2600	3100	3100	4100
Number of beams	426	122	40	277	80

The visual grading and strength class assignment were carried out according to the criteria of UNE-56544:1997 [4] standard because -as noted above- it includes *Populus* spp. timber, while in all new versions, it has been disregarded. The pieces were visually graded into three categories: two structural grades (ME-1, ME-2) and a non-structural grade (Rejection, R). The assignment to strength class for ME-1 was C18, for ME-2 was C14, and for Rejection was non-structural timber. Also, three sawing systems (radial, tangential and mixed) were considered in all specimens, which were indicated as “cut type”.

2.3. Non-destructive tests

Once the specimens were classified, a non-destructive characterization was carried out. Ultrasound (Sylvatest®) and vibrational analysis (FFT analyzer) methods were used to determine ultrasonic wave velocity and vibrational frequency, respectively. In both methods, the tests were carried out on the longitudinal direction of specimens (Figs. 4 and 5, respectively). The dynamic MOE was determined according to Eqs. (1) and (2) for ultrasound and vibrational analysis, respectively (NDT

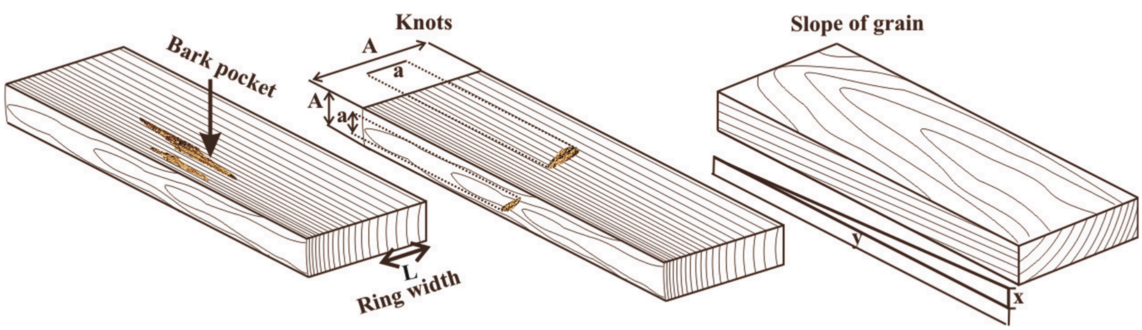

Fig. 1. Defects linked to timber anatomy.

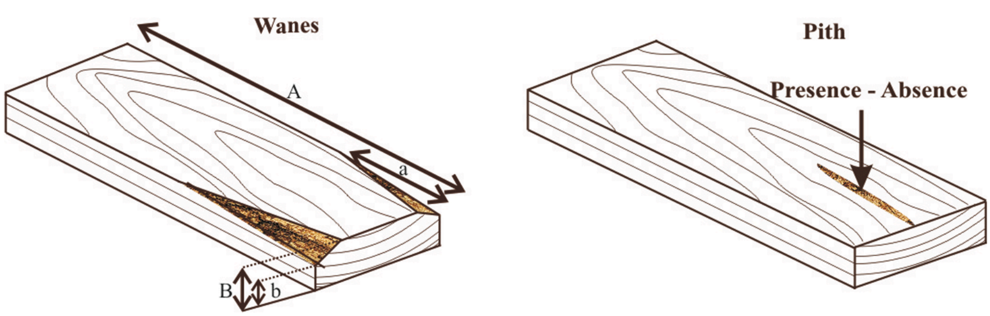

Fig. 2. Defects linked to sawing.

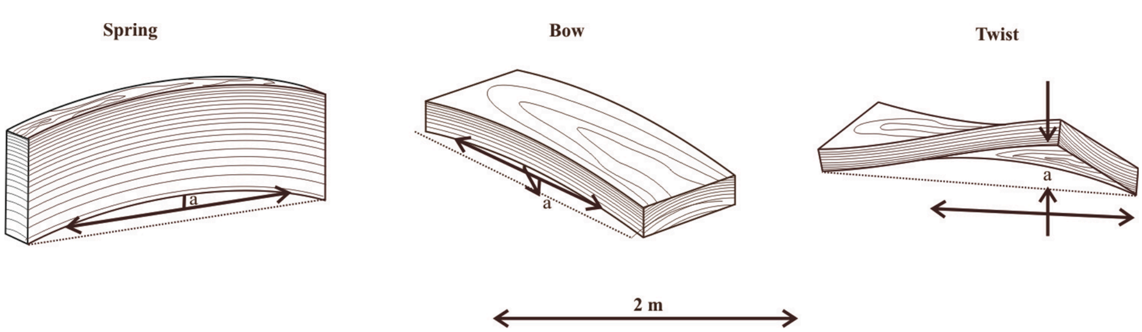

Fig. 3. Defects linked to the drying process.

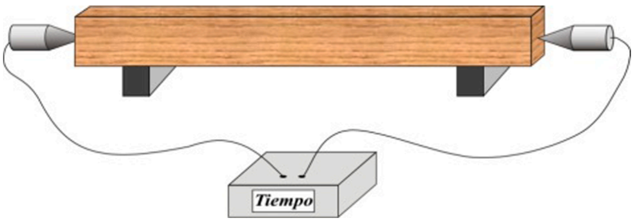

Fig.4. Ultrasonic method.

Fig.5. Vibrational analysis method.

variables).

$$MOE_{ult} = (V)^2 \times \rho \quad (1)$$

$$MOE_{vib} = (2 \times L \times f)^2 \times \rho \quad (2)$$

The recorded variables were: MOE_{ult}: modulus of elasticity obtained by ultrasonic method; V: ultrasonic wave velocity ($m \cdot s^{-1}$); ρ: density ($kg \cdot m^{-3}$), obtained by weight and volume determination; MOE_{vib}: modulus of elasticity obtained by vibrational analysis method; L: specimen length (m); f: vibrational frequency (Hz).

2.4. Physical-mechanical tests

Physical-mechanical characterization tests were performed following the methodology proposed in EN-408:2011 + A1:2012 [40]. These tests were conducted to obtain the physical-mechanical values and validate the strength classes assigned through visual grading. The static bending destructive test was conducted using a universal machine (ELIB-100 W-IBERTEST), with a 100 kN load cell and a central linear variable differential transformer (LVDT), as shown in Fig. 6. The stiffness and strength of static bending were determined through the global modulus of elasticity (MOE) and the modulus of rupture (MOR), respectively. Characteristic values were determined according to EN-384:2010 (taking the 5th percentile of density and MOR, and the mean of MOE) [10].

2.5. Statistical analysis

Data from 945 samples corresponding to the three different structural visual grades (ME-1, ME-2 and Rejection "R") were analyzed. First, the assumptions of independence, normality and homoscedasticity of the density, MOR and MOE data were checked for all groups. Data normality was checked for all populations using the Kolmogorov-Smirnov normality test with Lilliefors (L-KS) correction and further checked with the Normal Probability Plot. The homoscedasticity requirement was contrasted by the Levene test. In those cases in which the normality and homoscedasticity requirements were not met, preventing the use of ANOVA, two statistical technics were used: the Kruskal-Wallis test, when the distribution was not normal but the groups were homoscedastic; Welch's heteroscedastic F test with trimmed means and Winsorized variances when neither normality nor homoscedasticity could be assumed. This latter robust procedure tests the equality of means by substituting trimmed means and Winsorized variances for the usual means and variances [41,42]. Also, bootstrap methods were used to establish robust confidence intervals for location [43] and robust homogenous groups. All the statistical analyses were performed using R software (v. 3.6.1 [44]).

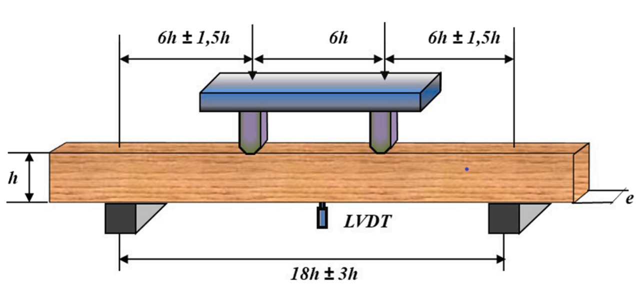

Fig. 6. Device and test conditions for the universal machine [40].

2.6. Classifying machine learning algorithms

A classification task using machine learning algorithms usually involves separating the data into training and test sets. The goal is to produce modelling based on the training data, able to predict the test data's target values, giving only the test data attributes. For grading structural timber, each instance (beam) in the training set contains one "target value" (i.e., the strength classes) and several "attributes" (observed or independent variables).

Once the characteristic values of all the beams were obtained, they were classified into resistance groups, defining three classes: class1 (C18), class2 (C14) and class3 (<C14 = reject). These were the target values and were actual strength values (i.e., obtained by destructive testing), not simply assigned (predicted) by visual grading.

A total of 9 ML algorithms were selected, trying to cover the most classical and commonly used algorithms and other modern and flexible algorithms that use packaging techniques (Table 2). Model tuning,

Table 2
Classifying machine learning algorithms, model type, basic paremeters used and references.

Algorithms	Model type	Basic parameters used	References
svm (Support Vector Machines)	Kernel methods	Kernel = radial basis funcion Kernel Cost parameter (C) = 1 gamma = auto optimized for 0.042388 number of support vectors = 679	[47]
nb (Naïve Bayes)	Probabilistic learning	Kernel density estimate for continuous variables was used. *Laplace smoother. fL = 0*	[48,49]
knn (K-Nearest Neighbor)	Lazy learning	9-nearest neighbor model	[50]
C5.0	Classification tree or rule-based models	Trials = 10 Model = rule	[51,52,53]
nnet	Neural networks	Multilayer percepton Size (Hidden units) = 1 Decay = 0.1	[54]
Rf (Random Forest)	Model ensembles – Decision trees	Number of trees: 500 mtry = 2	[55]
bagFDA (Bagged Flexible Discriminant Analysis)	Model ensembles Non-parametric multiple regression	degree = 1 nprune = 22	[56,57]
bagEarthGCV (Bagged MARS using gCV Pruning)	Multivariate adaptive Regression splines	degree = 1	[56]
Gbm (Gradient Boosting Machine)	Boosted trees	n.trees = 150 interaction.depth = 1 shrinkage = 0.1 n.minobsinnode = 10	[58–60]

Table 3
Variables used for each different scenarios.

Scenarios			Independent variables used (attributes)	Description (units of measurement)	Type	Min, Max values
Scenario 3: VAR3 = VAR2 + NDT variables	Scenario 2: VAR2 = VAR1 + "cut type"	Scenario 1: VAR1 = visual variables	V1	Face knots size (mm)	numerical	0 , 123.1
			V2	Face knots (absence/presence)	factor	0 , 1
			V3	Edge knots size (mm)	numerical	0, 192.3
			V4	Edge knots (absence/presence)	factor	0 , 1
			V5	Blue stain (absence/presence)	factor	0 , 1
			V6	Bark pockets (absence/presence)	factor	0 , 1
			V7	Pith (absence/presence)	factor	0 , 1
			V8	Bow (mm)	numerical	0 , 3.9
			V9	Spring (mm)	numerical	0 , 4.6
			V10	Twist (mm)	numerical	0 , 6.0
			V11	Slope of grain (mm)	numerical	0 , 58.4
			V12	Rings width (mm)	numerical	3.8 , 22.0
			V13	Wane width (mm)	numerical	0 , 20.0
			V14	Wane length (mm)	numerical	0 , 360.5
			V15	"cut type" (radial/tangential/mixed)	factor	R/T/M
			V17	MOE_{vib} (MPa)	numerical	5439 , 14799
			V18	MOE_{ult} (MPa)	numerical	4947 , 16321

training and prediction were performed using the caret package [45,46] as a wrapper package for an extensive list of machine learning algorithms implemented in R.

2.7. Input variables: analysis groups

The classification ML algorithms were applied in three scenarios, as follows: scenario 1: specimens grouped using visual (defects) variables (set of input variables VAR1); scenario 2: adding "cut type" (i.e. radial, tangential or mixed sawing systems) variable (set VAR2 = VAR1 + "cut type"); scenario 3: adding, at the same time, two NDT (MOEult and MOEvib) variables (set VAR3 = VAR2 + NDT variables); Table 3.

In order to sort the independent input variables according to their relative importance in the classification, a sensitivity analysis was performed.

2.8. Data preprocess

Data preprocessing is an essential task in data analysis [61], particularly when working with ML algorithms because many of them present a significant bias in their results when the input variables (attributes or independent variables) sizes are very different. In this study, it was observed that the numerical variables range was very changeable, finding extreme values that ranged from (0, 1) for "blue stain" to (4947, 16321) for MOE_{ult} (Table 3). This constitutes a problem in some models' final results that gives higher importance to variables with higher size values. Centring and scaling are the most usual data preprocess to achieve the same order of magnitude among variables. Centring removes the differences in the size of variables by subtracting the variable mean value from each element. Scaling allows for balancing the numerical values of the variables to achieve the same order of magnitude among variables, thus avoiding that attributes in greater numeric ranges dominate those in smaller numeric ranges while avoiding numerical difficulties during the calculation [62]. Moreover, feature value scaling can help to increase the model accuracy according to our experimental results. In this work, each variable was linearly scaled to the (0, 1) range, according to Eq. (3).

$$v' = \frac{v - min_a}{max_a - min_a} \tag{3}$$

where v′ is scaled value, v is the original value, min_a is the low bound of the feature value, and max_a is the upper bound of the feature value.

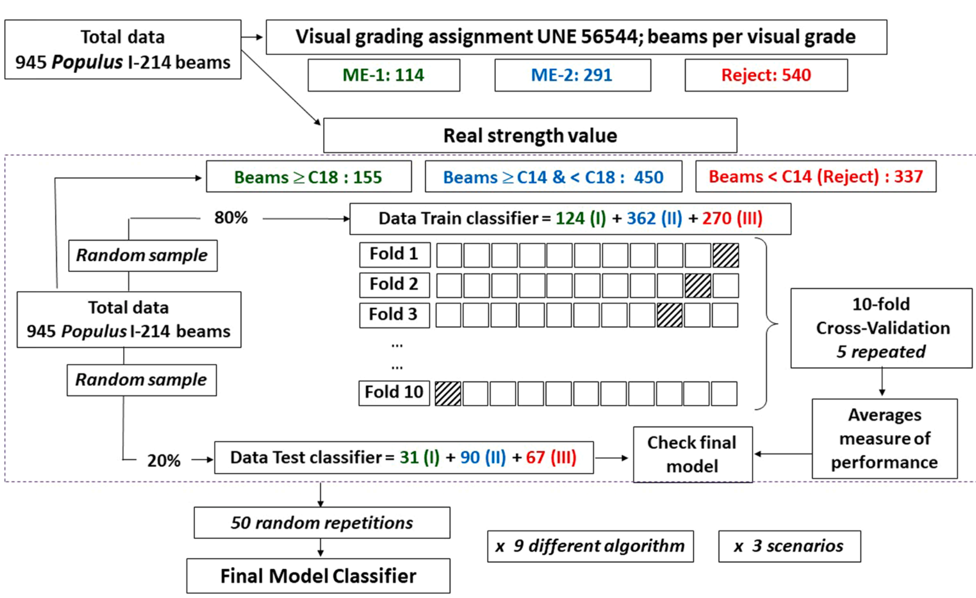

Fig.7. Outline of the methodological strategy to establish the machine learning algorithms.

The most classic ML algorithms require their input to be numerical; therefore, before using any of them, all categorical features were transformed into numerical features (dummy variables) that take the numeric values [63–65].

2.9. Model control: cross-validation

The goal of cross-validation is to test the model's ability to predict new data that were not used in estimating it [66]. Thus, it is a method to estimate the error rate efficiently and in an unbiased way.

This study used this technique in each of the algorithms, splitting the total pieces (945 beams) into a training set of 756 beams (80% of total) and a test set of 198 beams (20% of total). These percentages were selected to combine a good accuracy (better with high training percentage) and low overfitting (better with low training percentage) in all models. In this work, all algorithms used k-fold cross-validation (k-fold = 10), and the method was repeated n times (n = 5), yielding 50 different random partitions of the original sample. These 50 results were again averaged to produce a single estimation (Fig. 7). In this way, it was ensured that the statistical performance values found for each algorithm were highly robust.

2.10. Classification performance. Model evaluation metrics

Finally, a multi-class confusion matrix assessed the qualifier grading performance, estimating different statistics that allowed to determine the efficiency of the classificatory modelling methods. The classification problem consisted of three classes: C18 beams (or better), C14 beams, and reject. Fig. 8 shows the general confusion matrix (a) together with three matrices (b, c, d) in which the values TP (true positive), FP (false positive), TN (true negative) and FN (false negative) represent each of the classes.

The number of data belonging to each of the classes was not the same, being slightly unbalanced, and it was necessary to use metrics that take into account this problem. Branco *et al.* (2017) and Tharwat (2018) [67,68] reported some of the measures that derive from the multi-class confusion matrix for evaluating a diagnostic test. All these metrics have been proposed to assess the performance in multi-class imbalanced domains. In the present work, the performance of the metrics used is shown in Table 4.

For Global Metrics, Accuracy (*Acc*) and Kappa (*k*) are the main metrics used to evaluate algorithms classification problems. *Acc* is the percentage of correctly classified instances out of all instances. It is more useful on a binary classification than in multi-class classification because it can be less clear how the accuracy breaks down across those classes. To partially solve this problem, Overall Balanced Accuracy (*OBAcc*) [69] and Average Accuracy (*AvAcc*) are used to ponder the value of the *Acc* according to the weight of the classes and are commonly used in unbalanced multiclass classifications. Kappa or Cohen's Kappa (*k)* [70] is similar to *Acc*, but it is normalized at the baseline of random chance on the dataset. Thus, it is a more practical measure to use on problems with an imbalance in the classes. Landis and Koch (1977) [70] provide a way to characterize values. Matthews Correlation Coefficient (*MCC)* [71,72] and Confusion Entropy (*CEN*) [73,74] have been further developed and provide a more effective measure of grading performance in multi-class classifications. Finally, it is crucial to control the percentage of overgraded, as overgrading leads to unsafe and dangerous structures.

All these metrics are accompanied by a basic statistical analysis in evaluating the predictive capacity of any classification method, which is the no-information rate, i.e., the accuracy achievable by always predicting the majority (most common) class label.

Regarding individual classes, there are three well-known metrics: sensitivity, specificity and precision. They can estimate the classification performance achieved in each class and are calculated on each class by separately encoding different aspects of the classification. *Sensitivity* or *Recall* of the i^{th} class represents the ability of a given classifier to correctly identify the samples of the i^{th} class. *Precision* of the i^{th} class is defined as the purity of a class, that is, the classifier's ability to avoid wrong predictions in that class. *Specificity* of the i^{th} class represents the ability of a classifier to reject samples of other classes, and it is calculated as the ratio of samples not belonging to the i^{th} class which were not classified in the i^{th} class over the total number of samples not belonging to the i^{th} class. All of them, sensitivity, precision and specificity, have values between 0 (no class discrimination) and 1 (perfect class

a.- Omnibus		Real strength values		
Predicted Class		1	2	3
	1	A	B	C
	2	D	E	F
	3	G	H	I

b.- For Class 1		Real strength values		
Predicted Class		1	2	3
	1	A TP_1	B FP_1	C FP_1
	2	D FN_1	E TN_1	F TN_1
	3	G FN_1	H TN_1	I TN_1

c.- For Class 2		Real strength values		
Predicted Class		1	2	3
	1	A TN_2	B FN_2	C TN_2
	2	D FP_2	E TP_2	F FP_2
	3	G TN_2	H FN_2	I TN_2

d.- For Class 3		Real strength values		
Predicted Class		1	2	3
	1	A TN_3	B TN_3	C FN_3
	2	D TN_3	E TN_3	F FN_3
	3	G FP_3	H FP_3	I TP_3

Fig. 8. General confusion matrix (a) and performance values by class (b, c, d).

Table 4
Metrics for the evaluation of classification performance.

Metric	Description	Definition
Global metrics		
Acc	Accuracy (Overall accuracy)	$Acc = \frac{TP_1 + TP_2 + TP_3}{\text{Total data}}$
OBAcc	Overall Balanced Accuracy	$OBAcc = \left[\frac{TP_1}{TP_1 + \sum FP_1}\right] + \left[\frac{TP_2}{TP_2 + \sum FP_2}\right] + \left[\frac{TP_3}{TP_3 + \sum FP_3}\right] \Big/ 3$
AvAcc	Average Accuracy	$AvAcc = \left[\frac{TP}{TP + FP_1 + FN_1}\right] + \left[\frac{TP}{TP + FP_2 + FN_2}\right] + \left[\frac{TP}{TP + FP_3 + FN_3}\right] \Big/ 3$ where $TP = \Sigma TP_i$
k	Kappa or Cohen's Kappa	$k = \frac{p_0 - p_e}{1 - p_e}$ where: p_o = observed agreement; p_e = expected agreement
MCC	Matthews Correlation Coef. extended to multi-class	$\frac{X}{YZ}$; $X = \left[\sum_{k,l,m=1}^{C} (mat_{k,k}mat_{m,l} - mat_{l,k}mat_{k,m})\right]$ $Y = \sqrt{\sum_{k=1}^{C}\left(\sum_{k=1}^{C} mat_{l,k}\right)\left(\sum_{f,g=1,\ f \neq k}^{C} mat_{g,f}\right)}$ $Z = \sqrt{\sum_{k=1}^{C}\left(\sum_{l=1}^{C} mat_{k,l}\right)\left(\sum_{f,g=1,\ f \neq k}^{C} mat_{f,g}\right)}$
CEN	Confusion Entropy	$CEN = \sum_{j=1}^{C} (P_j CEN_j); P_j = \frac{\sum_{k=1}^{C} mat_{j,k} + mat_{k,j}}{2*\sum_{k,l=1}^{C} mat_{k,l}}$ $CEN_j = \sum_{k=1,\ k \neq j}^{C} (P_{j,k}^{j} log_{2(C-1)}(P_{j,k}^{j}) + P_{k,j}^{j} log_{2(C-1)}(P_{k,j}^{j}))$
Overgraded	Percentage of assigned upgrades	$Overgrades = \frac{B + C + F}{Total\ data} * 100$
Metrics by Class		
Sen = Sensitivity or Recall	Sensitivity by class	$Sensitivity_{class\ i} = \frac{TP_i}{TP_i + FN_i}$
Spe = Specificity	Specificity by class	$Specifity_{class\ i} = \frac{TN_i}{TN_i + TP_i}$
Precision	Precision by class	$Precision_{class\ i} = \frac{TP_i}{TP_i + FP_i}$
Prev = Prevalence	Prevalence by class	$Prevalence_{class\ i} = \frac{TP_i + FN_i}{\text{Total data}}$
NPV or TNA	Negative predictive value or true negative accuracy	$NPD_{class\ i} = \frac{Sen_i * Prev_i}{(Sen_i * Prev_i) + ((1 - Spe_i) * (1 - Prev_i))}$
Detection rate	True positive rate	$TPR_{class\ i} = \frac{TP_i}{\text{Total data}}$
Detection prevalence	Detection prevalence	$TPR_{class\ i} = \frac{TP_i + FP_i}{\text{Total data}}$
F1	F1 by class	$F1\ score_{class\ i} = \frac{2*TN_i}{TP_i + FP_i + FN_i}$
Balanced Acc	Balanced accuracy by class	$Balanced\ Acc = \frac{Sensivity_i + Specificity_i}{2}$
CEN_j	Confusion entropy by class	$CEN_{classj} = \sum_{k=1,\ k \neq j}^{C} (P_{j,k}^{j} log_{2(C-1)}(P_{j,k}^{j}) + P_{k,j}^{j} log_{2(C-1)}(P_{k,j}^{j}))$

discrimination) [75]. Balanced Accuracy is essentially an average of Sensitivity and Precision, and the main difference between Balanced Accuracy and Accuracy emerges when the initial set of data shows an unbalanced distribution for the classes.

In addition to the above metrics, *Prevalence, NPV, Detection rate, Detection prevalence* and *F1* were also used. *Prevalence* shows how often the positive class actually occurs in our sample. *NPV* measures the proportion of negative samples correctly classified to the total number of negative predicted samples. *Detection Rate* shows the number of correct positive class predictions made as a proportion of all of the predictions made, and finally, *Detection Prevalence* shows the number of positive class predictions made as a proportion of all predictions. These three metrics try to estimate the relevance of a class. Also, *F1* assesses the classification model's performance starting from the confusion matrix, aggregating *Precision* and *Sensitivity* measures under the concept of harmonic mean.

3. Results and discussion

3.1. Visual strength grading

Frequency histograms and boxplots for the three main physical-mechanical parameters (MOE, MOR and density) in each visual grade are shown in Fig. 9.

The main descriptive statistics of MOE, MOR and density values, by visual strength grades, are shown in Table 5.

According to Fig. 9 and Table 5, the assumption of normality and homoscedasticity of the samples is not fulfilled in most groups.

Among the different visual strength grades, no significant differences were observed for density, although the mean value of MOE (8657.9 MPa) and MOR (44.3 MPa) of ME-1 grade were higher and statistically different than those of ME-2 and Rejection grades. The key values of the latter two grades did not show significant differences.

Regarding the characteristic values of each strength class, the ME-1

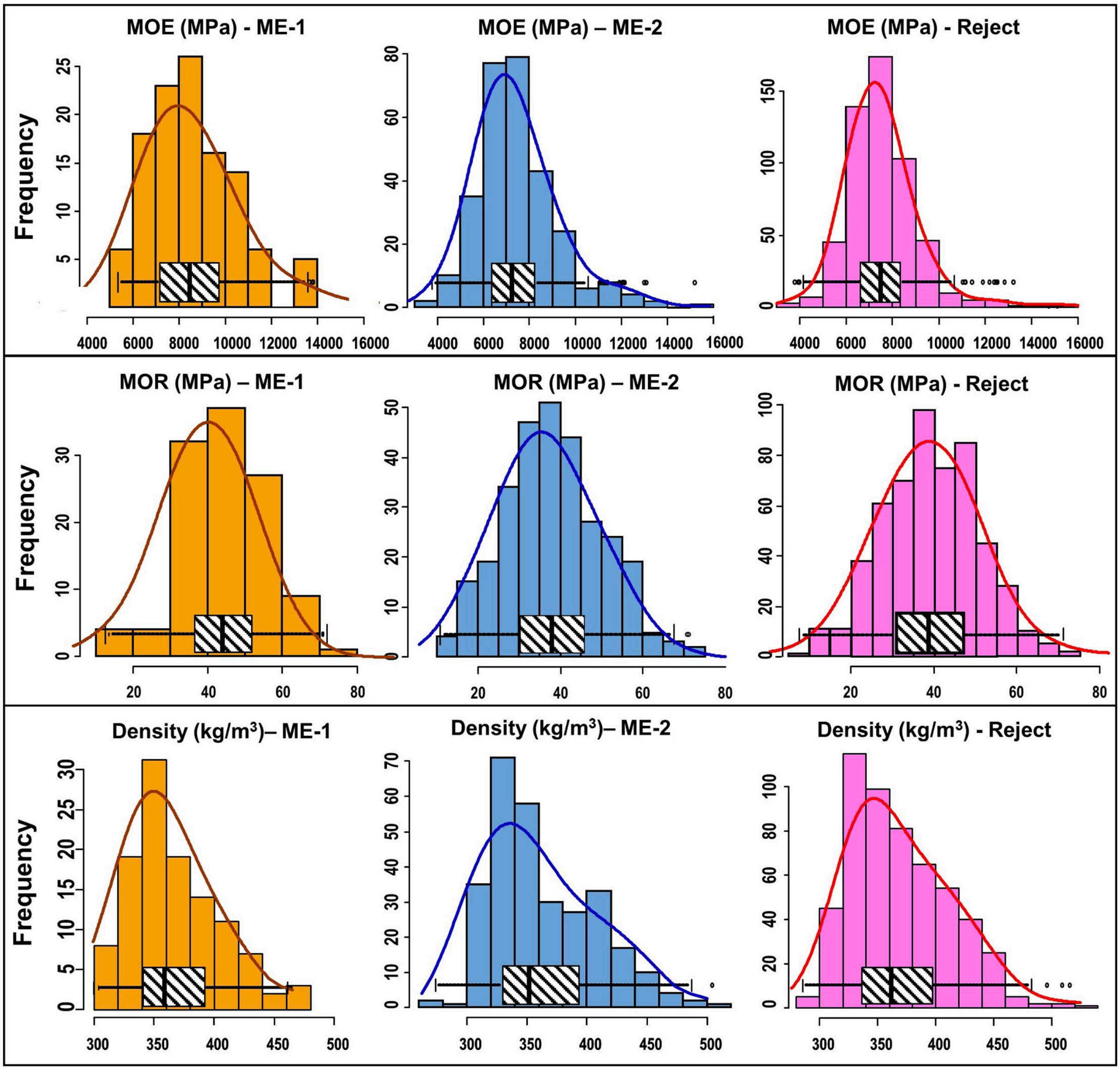

Fig. 9. Frequency histograms and boxplots for MOE, MOR and density.

grade did not comply with the minimum values established in standard EN-338 for C18, which requires a MOE $\geq$ 9000 MPa, and, taking into account the characteristic values determined in the tests, their allocation should be set to C16.

The classifying performance of the visual grading was established using the confusion matrix, together with the Global metrics and Metrics by class mentioned above (Table 6).

The accuracy/overall accuracy (Acc = 37.78% with a 95% CI from 34.68% to 40.96%) and unweighted Kappa statistic (Kappa = 0.011) showed no reliability, and the other global metrics to check the performance of visual grading corroborated this statement.

Particular emphasis should be placed on the values of the no-information rate and overgrades.

The no-information rate (NIR = 47.9%) is the accuracy achievable by always predicting the most common class label in the test set. A one-sided hypothesis test was calculated to assess whether the overall accuracy rate is greater than NIR (p-value = 1), i.e., if by assigning all the beams to majority class (ME-2), we would have an accuracy equal to or greater than that obtained by applying the laborious visual grading, and it suggests that the visual rule needs to be revised for this species.

About overgrades, a high number of them pose a serious safety problem, as higher characteristic values than those that the beam has being assigned. The overgrades value obtained using the visual classification was 20.53 %.

According to the percentage of results produced using visual grading standards, a poor level of efficacy UNE-56544:1997 can be inferred,

Table 5
Statistical summary of MOE, MOR and density by visual strength grades.

	Visual grades	n	Mean ± robust CI (rob. homog. groups)*	5th Percentile	p-value L-KS	p-value Levenés Test Test between groups (p-value)
MOE (MPa)	ME-1	114	8658 ± 343 (a)	6090.9	0.029	3.4e-04
	ME-2	291	7440 ± 192 (b)	5287.6	4.9e−07	Welchś test (0.00)
	Reject	540	7566 ± 121 (b)	5662.4	1.0e−07	
MOR (MPa)	ME-1	114	44.3 ± 2.0 (a)	26.4	0.750	0.880
	ME-2	291	38.1 ± 1.3 (b)	19.1	0.510	ANOVA (5.4e-06)
	Reject	540	39 ± 1.0 (b)	20.6	0.656	
Density (kg/m3)	ME-1	114	368 ± 6.9 (a)	316.9	0.002	0.223
	ME-2	291	363 ± 4.9 (a)	313.5	1.2e−07	Kruskall-Wallis (0.081)
	Reject	540	369 ± 3.6 (a)	314.7	2.6e−10	

*Different letters (between parentheses) indicate significative differences between visual strength grades for each parameter.

Table 6
Confusion matrix and global metrics of visual grading.

		Confusion Matrix Reference classes			
		C18	C14	Reject	
Prediction classes	ME-1	41	49	24	In green = number of correct values
	ME-2	46	124	121	In orange = number of overgrades
	Reject	68	280	192	In white = number of undergades

Global metrics		Metrics by class	ME-1	ME-2	Reject
Accuracy:	**0.378**				
95% CI:	(0.347, 0.410)	Sensitivity	0.265	0.274	0.570
No Information Rate (NIR):	**0.479**	Specificity	0.908	0.661	0.428
P-Value [Acc > NIR]:	1	Precision	0.360	0.426	0.356
Kappa:	0.011	F1 score	0.305	0.333	0.642
Mcnemar's Test P-Value:	<2e-16	Neg Pred Value	0.863	0.497	0.642
Overgrades:	20.53%	Prevalence	0.164	0.479	0.357
Overall Balance Accuracy:	0.369	Detection Rate	0.043	0.131	0.203
Average Accuracy (AvACC):	0.498	Detection Prevalence	0.121	0.308	0.571
Confusion Entropy:	0.816	Balanced Accuracy	0.586	0.467	0.499
Matthews Corr. Coef. (MCC):	0.011	Confusion Entropy	0.551	0.954	0.944

with an acceptable accuracy mainly obtained in reject class and an important percentage of overgrades. There was an apparent mismatch between the strength values required by the UNE-EN-338 standard and the actual values observed in the species. Concerning the high number of over-graded beams, as mentioned by [76], in the under-graded or over-graded results, the business consequences are different: the under-graded timber leads to the underestimation of the price of the piece of wood, thus the people who made the grading lose money; the over-graded timber is dangerous because it will be used at loads it cannot stand and, as a consequence, buildings can be damaged or even collapse.

Similar results concerning grading inefficiency of UNE-56544:1997 [4] standards were previously observed by [17] for *Populus* x *euramericana* I-214 structural timber classification. In this case, the standard resulted in an excessively high rejection rate, with a consequent reduction in the value of timber batches. The performance of visual grading in [17] provided a high accuracy rate, with many rejected and under-graded pieces. Such research proposed an innovative structural grading standard for this timber with two strength classes (CHP2 and CHP1) that have less demanding MOE, MOR and density values (CHP2: MOE: 6500 N·mm^{-2}, MOR: 18 N·mm^{-2} and density: 290 kg·m^{-3}; CHP1: MOE 8000 N·mm^{-2}, MOR: 22 N·mm^{-2} and density: 310 kg·m^{-3}), primarily to provide a disqualifying effect for high elasticity values, which is required in EN-338:2010 [9]. The implementation of these strength classes and "reclassifying" the analyzed timber improved their results.

3.2. Machine learning classifiers

Nine different classifiers were used. Fifty random models were generated, supported by repeated cross-validation (k-fold = 10 and 5 times repeated) for each of them. Fig. 10 shows the global metrics for the three scenarios described in Table 2.

For all nine algorithms used in the three scenarios, the values obtained in all the global metrics were significantly better than those obtained in the visual grading methodology. Further, in all algorithms, the accuracy was remarkably better than NIR, and the values of scenario 3 were always better than those in the other two cases.

The numerical values of the accuracy varied between 0.515 (nb) and 0.548 (gbm) in scenario 1 (Appendix A); between 0.528 (nb) and 0.574 (rf) in scenario 2 (Appendix B); and between 0.603 (nb) and 0.661 (rf) in scenario 3 (Appendix C), in comparison with a 0.378 value for the visual grading. Average Accuracy (AvACC) varied between 0.626 (nb) and 0.665 (svm) in scenario 1 (Appendix A); between 0.638 (nb) and 0.680 (rf) in scenario 2 (Appendix B); and between 0.703 (nb) and 0.751 (rf) in scenario 3 (Appendix C), versus 0.498 for the visual grading. With respect to overgrades, they varied between 12.6% (nb) and 22.6% (C5.0) in scenario 1 (Appendix A); between 13.7% (nb) and 22.3% (nnet) in scenario 2 (Appendix B); and between 12.0% (nb) and 20.3% (knn) in scenario 3 (Appendix C), versus 20.53% for the visual grading.

Concerning the rest of the global metrics, the behaviour was the same

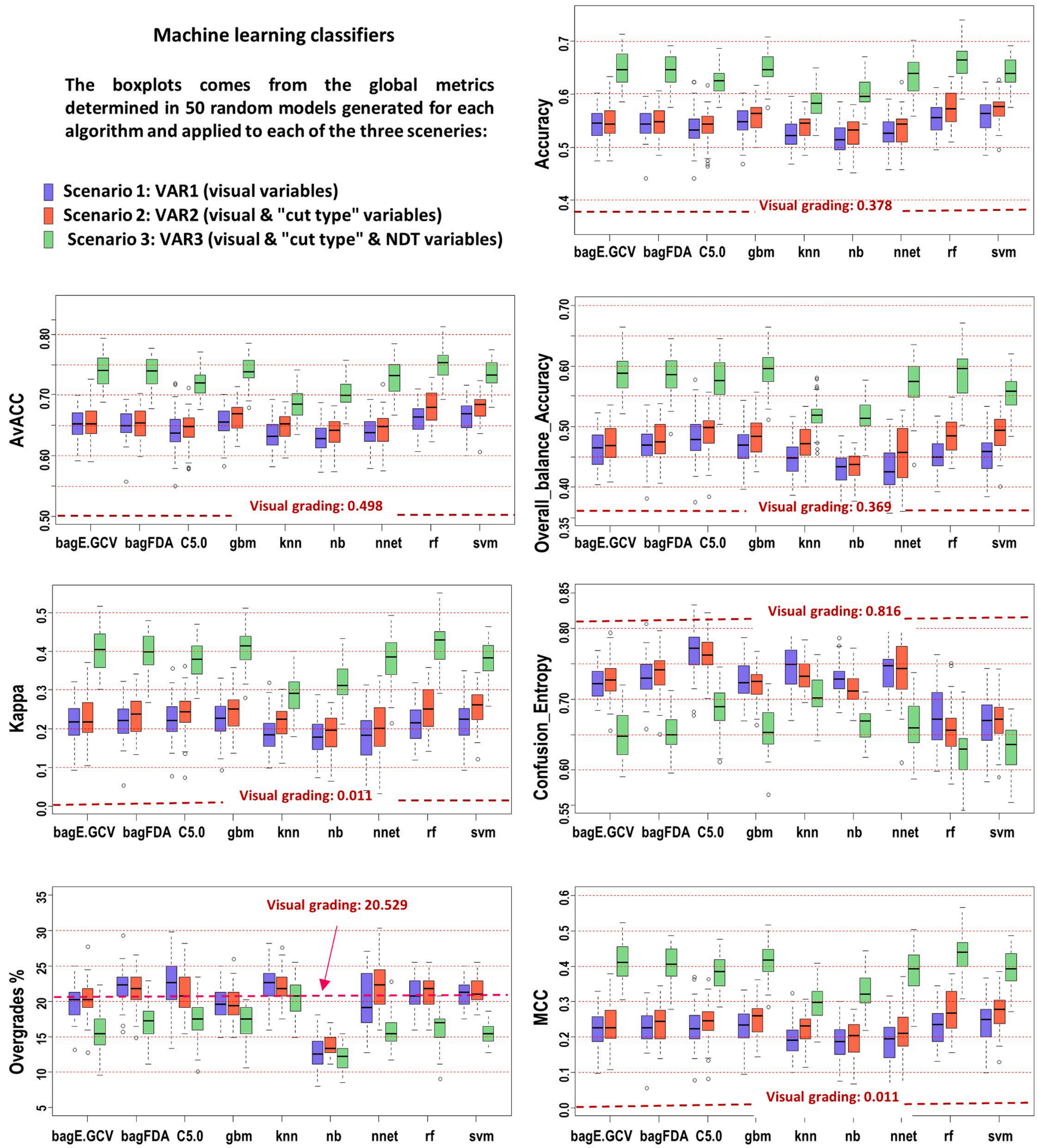

Fig. 10. Global metrics for different ML algorithms and scenarios.

as in the cases mentioned above. For any algorithm in any of the three scenarios, the values of the classification metrics were clearly better than those obtained by visual grading.

As for the metrics by class, the same response as in the global metrics was found (Fig. 11; Appendix D, scenario 1; Appendix E, scenario 2; Appendix F, scenario 3).

A better response was always obtained in the ML classifications than in the visual grading, finding the best values for scenario 3, in which the non-destructive variables, MOE_{vib} and MOE_{ult}, were incorporated (Fig. 11). These variables, when independently used, had a poor classification performance for *Populus* I-214 timber, offering accuracy values of 30% for MOE_{ult} and 46% for MOE_{vib} (Appendix G Supplementary figure 1), so it was decided to use them as a means of improving the overall performance of the classification algorithms. In this sense, [17] improved the prediction of stiffness-strength variables when they combined visual grading and NDT variables, although the methodology used in that implementation was merely additive. In this respect, there are many works about timber from different Spanish species (hardwoods and softwoods) that have found inefficient grading in some visual grading standards, and some of them also chose to include NDT variables to improve the classification or to predict the mechanical values [77–84, among others].

The influence of sawing systems ("cut type": radial, tangential or mixed) variable was mainly related to wane size (width and length). In this regard, a variation on sections and symmetries of beams, by wane presence, is linked to construction issues, such as the convenience of a flat surface for easier use (support pieces, nailing and glueing, etc.); but, in general, the resulting loss in strength capacity for a timber piece is

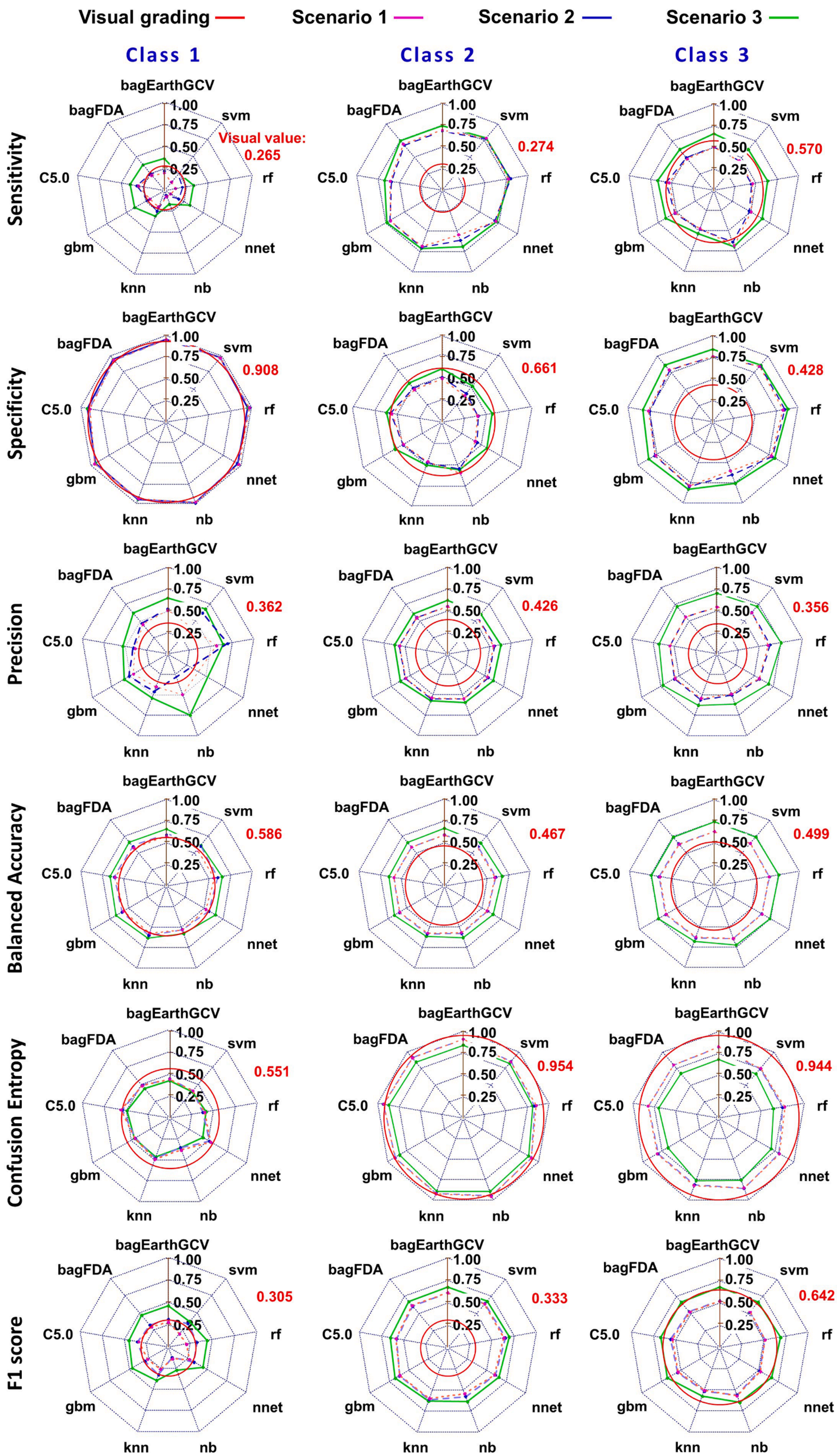

Fig. 11. Metrics by class for the different ML algorithms and scenarios.

secondary to the highlighted construction issues [85]. Moreover, the influence of warps may be secondary to the loss in strength capacity too, but sometimes they are related to the presence of pith or with ring width and should therefore be considered. Diez and Fernández-Golfín (1998) [86] studied the influence of various factors on *Pinus sylvestris* structural timber grade, including the influence of the sawing system. They found that improved sawing systems could reduce grade variability of timber pieces and increase physical-mechanical properties, albeit at the expense of a sawing performance loss. Following the above discussion, NDTs and sawing variables could be included as predictor variables in grading criteria.

Finally, the characteristic values of the grading groups obtained with the algorithms were checked. The random forest algorithm (Rf) is shown as an example, not because it was better or worse than the rest, but because it is one of the most traditional, known and studied algorithms in biosciences. The results are shown in Table 7, where it is highlighted in bold letters that, for the three scenarios, the groups met the requirements of the standard for strength classes [9].

The Rf classifications were better than those obtained with visual grading. For example, in scenario 1, with the same visual variables used, using the rf algorithm on the analyzed dataset led to the achievement of EN-338 requirements. The classification values of the other algorithms

Table 7
Statistical summary of MOE, MOR and density for the three scenarios using the rf algorithm.

	rf classes	n	Mean ± robust IC (rob. homog. groups)*		5th Percen.	p-value L-KS	p-value Levene Test between groups (p-value)
Scenario 1							
MOE	Class 1	14	**8995** ± 808	(a)	7012.1	0.812	0.161
(MPa)	Class 2	114	**8002** ± 171	(b)	6193.9	0.124	ANOVA (2.7e-10)
	Class 3 (Reject)	60	**6678** ± 291	(c)	4972.9	0.316	
MOR	Class 1	14	45.2 ± 4.6	(a)	**30.0**	0.695	0.482
(MPa)	Class 2	114	40.4 ± 2.0	(b)	**22.4**	0.394	ANOVA (2.2e-03)
	Class 3 (Reject)	60	35.3 ± 2.9	(c)	**18.6**	0.794	
Density	Class 1	14	366.0 ± 22.0	(a)	**318.0**	0.137	0.750
(kg/m^3)	Class 2	114	366.5 ± 8.3	(a)	**317.2**	2.97 e-06	Kruskall-Wallis (0.4)
	Class 3 (Reject)	60	357.2 ± 9.7	(a)	**313.5**	4.64 e-04	
Scenario 2							
MOE	Class 1	13	**10110** ± 1254	(a)	6962.9	0.3108	0.001
(MPa)	Class 2	136	**7761** ± 239	(b)	5880.0	2.17 e-4	Welch's test (1.4e-06)
	Class 3 (Reject)	39	**6537** ± 389	(c)	4416.5	0.3113	
MOR	Class 1	13	48.7 ± 6.3	(a)	**30.0**	0.182	0.767
(MPa)	Class 2	136	40.9 ± 1.9	(b)	**22.4**	0.260	ANOVA (2.1 e-03)
	Reject	39	35.8 ± 3.8	(c)	**18.6**	0.569	
Density	Class 1	13	409.6 ± 27.0	(a)	**336.9**	0.635	0.213
(kg/m^3)	Class 2	136	368.6 ± 6.9	(b)	**316.6**	4.44 e-04	Kruskall-Wallis (2.8 e-03)
	Class 3 (Reject)	39	357.3 ± 12.0	(b)	**313.9**	1.90 e-03	
Scenario 3							
MOE	Class 1	24	**9787** ± 465	(a)	6962.9	0.614	0.008
(MPa)	Class 2	110	**7773** ± 111	(b)	5880.0	0.643	Welch's test (<2 e-16)
	Class 3 (Reject)	54	**6460** ± 152	(c)	4416.5	0.112	
MOR	Class 1	24	41.68 ± 4.6	(a)	**30.0**	0.380	0.980
(MPa)	Class 2	110	40.99 ± 1.9	(b)	**22.4**	0.078	ANOVA (7.5 e-4)
	Class 3 (Reject)	54	34.26 ± 2.3	(c)	**18.6**	0.470	
Density	Class 1	24	393.5 ± 18.0	(a)	**336.9**	0.066	0.035
(kg/m^3)	Class 2	110	364.4 ± 8.8	(b)	**316.6**	1.6 e-07	Welch's test (4.9 e-05)
	Class 3 (Reject)	54	348.3 ± 8.5	(c)	**313.9**	1.3 e-05	

*Different letters (between parentheses) indicate significative differences between rf strength classes for each parameter. The characteristic values appear in bold letters.

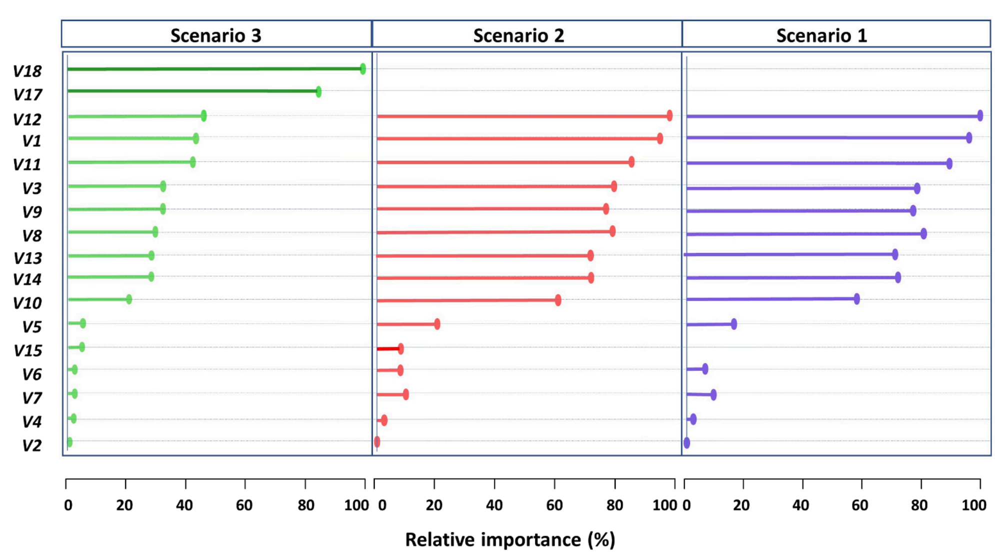

Fig. 12. Relative importance in Rf algorithm for the three scenarios.

used in this work, in the three scenarios, are shown in Appendix H; and all of them fulfil the EN-338 requirements [9].

Fig. 12 shows, for all three scenarios, the relative importance of independent input variables in output response, which was estimated through a sensitivity analysis after rf supervised learning model was built.

This sensitivity analysis corroborates the influence and improvement of timber grading through the inclusion of NDT variables in the model [81,82,84, among others]. In this sense, V17 (MOE_{vib}) and V18 (MOE_{ult}) added to the model, improved the accuracy from 55.5% in Scenario 1 -only visual variables- to 66.1% in Scenario 3 -visual, “cut type” and NDT variables- (Fig. 10; Appendix A; Appendix B). The most influential variables in Scenario 1 and 2 were V1 (Face knots size), V11 (Slope of grain) and V12 (Rings width). The presence of these defects may be related to the growth rate of the species. That is, poplar is a fast-growing species and, as such, growth and site conditions may influence the appearance of defects such as rings width, slope of grain, and even knots. Also, the presence and size of the latter may be related to the silvicultural treatments applied (or not) during its development (although this variable has not been into account in this study) [86,87]. Likewise, although variable V15 (“cut type”) was slightly influential, it significantly improved accuracy in scenarios 2 and 3.

4. Conclusions

According to the qualifying efficiency results in Spanish *Populus* x *euramericana* I-214 structural timber, a poor level of efficacy assignment of UNE-56544:1997 standard may be inferred, with acceptable accuracies mainly obtained in the 'reject' grade and with an important percentage of under and overgrades. The structural strength classes assigned to visual grades are limited in terms of mechanical values prediction effectiveness, with the economic and structural safety consequences that this may have. An apparent mismatch between the strength values required by the EN-338 standard and the observed values (“real”) was found, with an overgrading of many beams. The ME-1 grade did not comply with the minimum values established in standard EN-338 (C18 for ME-1), and taking into account the results, their allocation should be set to C16.

According to the global metrics and the metrics by class, all the Supervised Machine Learning algorithms showed a better classification performance than visual grading in the three scenarios, particularly in scenario 3. In this sense, NDTs variables combined with visual defects and sawing systems (“cut type”) were found to be better predictors than visual grading variables alone. These observations suggest that it is essential to consider NDTs and the effect of the sawing system as variables in timber grading criteria.

In response to the lack of structural visual strength grading for *Populus* x *euramericana* I-214 timber from Spain, Supervised Machine Learning algorithms could be an efficient and cost-effective tool to implement in the qualifying process.

Funding

This work was supported by the project of classification by non-destructive and pseudo-non-destructive methods and *Populus* structural wood, Junta Castilla y León, España [VA047A08].

CRediT authorship contribution statement

Luis Acuña-Rello: Conceptualization, Validation, Formal analysis, Methodology, Investigation, Writing – original draft, Writing – review & editing, Supervision. **Eleana Spavento:** Conceptualization, Validation, Formal analysis, Methodology, Investigation, Writing – original draft, Writing – review & editing, Supervision. **Milagros Casado-Sanz:** Methodology, Investigation, Writing – review & editing, Funding acquisition. **Luis-Alfonso Basterra:** Methodology, Investigation, Writing – review & editing. **Gamaliel López-Rodríguez:** Methodology, Investigation, Writing – review & editing. **Gemma Ramón-Cueto:** Methodology, Investigation, Writing – review & editing. **Enrique Relea-Gangas:** Methodology, Writing – review & editing. **Leandro Morillas-Romero:** Investigation, Writing – review & editing. **David Escolano-Margarit:** Investigation, Writing – review & editing. **Roberto D. Martínez:** Investigation, Writing – review & editing. **José Antonio Balmori:** Investigation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendices. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.engstruct.2021.113826.

References

[1] Ridley-Ellis D, Stapel P, Baño V. Strength grading of sawn timber in Europe: an explanation for engineers and researchers. Eur J Wood Prod 2016;74(3):291–306. https://doi.org/10.1007/s00107-016-1034-1.
[2] Stapel P, van de Kuilen J-W. Efficiency of visual strength grading of timber with respect to origin, species, cross section, and grading rules: a critical evaluation of the common standards. Holzforschung 2014;68(2):203–16. https://doi.org/10.1515/hf-2013-0042.
[3] Vega A, Arriaga F, Guaita M, Baño V. Proposal for visual grading criteria of structural timber of sweet chestnut from SpainVorschlag für die visuelle Sortierung von Bauholz aus spanischer Edelkastanie. Eur J Wood Prod 2013;71(4):529–32. https://doi.org/10.1007/s00107-013-0705-4.
[4] UNE 56544. Visual grading for structural sawn timber; 1997. p. 16.
[5] UNE 56544. Visual grading for structural sawn timber. Coniferous timber; 2011. p. 21.
[6] Glass S, Zelinka S. Moisture relations and physical properties of wood. In: USDA Centennial, editor. Wood Handbook: wood as an engineering material. Madison, Wisconsin: Forest Products Laboratory; 2010. p. 4.1–4.19.
[7] Vera Sepúlveda J, Ananías R, Díaz-Vaz J. Condicionantes de las grietas internas dentro de los anillos durante el secado de pino radiata [Some factors affecting the whiting-ring internal checking during drying of radiata pine]. Maderas Ciencia y tecnología 2007;9(2):145–59. https://doi.org/10.4067/S0718-221X2007000200005.
[8] Álvares NH. Deformaciones de la madera aserrada producidas durante y despues del secado. Hojas divulgativas 8/86HD [Warps of sawing timber produced before and after drying; Informative sheets 8/86HD]. Ministerio de Agricultura Pesca y Alimentacion; 1986.
[9] EN-338. Structural timber. Strength classes; 2010. p. 12.
[10] EN-384. Structural timber. Determination of characteristic values of mechanical properties and density; 2010. p. 20.
[11] EN-1912. Structural timber. Strength classes. Assignment of visual grades and species; 1999. p. 18.
[12] Llana DF, Íñiguez-González G, Díez MR, Arriaga F. Nondestructive testing used on timber in Spain: a literature review. Maderas-Cienc Tecnol 2020;22(2):133–56. https://doi.org/10.4067/S0718-221X2020005000201.
[13] Hermoso E. Caracterización mecánica de la madera estructural de *Pinus sylvestris* L. Tesis doctoral. [Mechanical characterization of structural timber of *Pinus sylvestris* L.; PhD thesis]. España: Universidad Politécnica de Madrid; 2001. 253 pp.
[14] Feio A, Machado JS. In-situ assessment of timber structural members: Combining information from visual strength grading and NDT/SDT methods-A review. Constr Build Mater 2015;101:1157–65. https://doi.org/10.1016/j.conbuildmat.2015.05.123.
[15] Barriola MJ, Aira JR, Villanueva JL. Analytical models of the mechanical properties of Japanese larch (Larix kaempferi(Lamb.) Carr.) based on non-destructive testing and visual grading parameters. Wood Mat Sci Eng 2021;16(2):94–101. https://doi.org/10.1080/17480272.2019.1626481.
[16] Fernández-Golfín Seco JI, Díez Barra MR, Hermoso PE. Análisis y estado del arte de la técnica de clasificación mecánica por máquina de la madera aserrada estructural (Revisión). [State of the art of the machine strength rating of lumber (Review)]. Invest Agr: Sist Recur For 2001;10(1):15.
[17] Casado M, Acuña L, Basterra LA, Cueto GR, Vecilla D. Grading of structural timber of *Populus* x *euramericana* clone I-214. Holzforschung 2012;66:633–8. https://doi.org/10.1515/hf-2011-0153.
[18] Arriaga F, Monton J, Segues E, Íñiguez-González G. Determination of the mechanical properties of radiata pine timber by means of longitudinal and transverse vibration methods. Holzforschung 2014;68(3):299–305. https://doi.org/10.1515/hf-2013-0087.

[19] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature 2015;521(7553):452–9. https://doi.org/10.1038/nature14541.
[20] Muhammad Redzuan FI, Yusoff M. Knots timber detection and classification with C-Support Vector Machine. Bull Electr Eng Informat 2019;8(1):246–52. https://doi.org/10.11591/eei.v8i1.1444.
[21] Sun YD, He JY, Wu MH, Zheng JJ, Gao Y, Wang XS. Study on Support Vector Machine Combined with Infrared Spectroscopy for Timber Species Identification; 2014. [International Conference on Future Computer and Communication Engineering].
[22] dos Santos JX, Vieira HC, Souza DV, de Menezes MC, de Muñiz GIB, Soffiatti P, et al. Discrimination of "Louros" wood from the Brazilian Amazon by near-infrared spectroscopy and machine learning techniques. Eur J Wood Wood Prod 2021;79 (4):989–98. https://doi.org/10.1007/s00107-021-01685-3.
[23] Ramírez Alonso GMJ, Chacón Murguía MI. Clasificación de defectos en madera utilizando Redes Neurales Artificiales [Wood defects classification using Artificial Neural Network]. Computación y Sistemas 2005;9(1):017–27.
[24] Tiryaki S, Hamzacebi C. Predicting modulus of rupture (MOR) and modulus of elasticity (MOE) of heat treated woods by artificial neural networks. Measurement 2014;49:266–74. https://doi.org/10.1016/j.measurement.2013.12.004.
[25] Tiryaki S, Aydin A. An artificial neural network model for predicting compression strength of heat treated woods and comparison with a multiple linear regression model. Constr Build Mater 2014;62:102–8. https://doi.org/10.1016/j.conbuildmat.2014.03.041.
[26] Bardak S, Tiryaki S, Nemli G, Aydın A. Investigation and neural network prediction of wood bonding quality based on pressing conditions. Int J Adhes Adhes 2016;68: 115–23. https://doi.org/10.1016/j.ijadhadh.2016.02.010.
[27] Pereira Miguel E, de Melo RR, Serenini Junior L, Soares Del Menezzi CH. Using artificial neural networks in estimating wood resistance. Maderas Ciencia y tecnología 2018;20(4):531–42. https://doi.org/10.4067/S0718-221X2018005004101.
[28] Mier Pérez R, García de Ceca JL, Díez Barra MR, Fernández-Golfín Seco JI, Hermoso Prieto E. Aplicación de redes neuronales a la clasificación de madera estructural. Comparación con otros métodos de clasificación. [IV Congreso Forestal Español. Sección 04]; 2005.
[29] ESYRCE. Encuesta sobre Superficies y Rendimientos de Cultivos [Survey on Surfaces and Crops Yields]. [En línea] 2019. [Citado el: 21 de May de 2020.] https://www.mapa.gob.es/es/estadistica/temas/estadisticas-agrarias/agricultura/esyrce/.
[30] Borodowski E. Situación actual del cultivo y uso de las Salicáceas en Argentina [Current situation of the cultivation and use of "Salicáceas" in Argentina]. Jornadas de Salicáceas-V Congreso Internacional de Salicáceas 2017:20.
[31] Thomas E, Garces A. Crecimiento inicial de 13 clones de álamo en el Valle Medio del río Negro. Comunicación [Initial growth of 13 poplar clones in the "Valle medio de Río Negro (Argentina)"; Communication]. III Congreso Internacional de Salicáceas, Argentina 2011:5.
[32] Castro G, Paganini G. New technologies and alternative uses for Poplar wood. [Boletín del CIDEU], vol. 2; 2009. p. 27–36. ISSN 1885-5237.
[33] Heräjärvi H. Effect of drying technology on Aspen wood properties. Silva Fennica 2009;43(3):433–45. https://doi.org/10.14214/sf.198.
[34] Baonza Merino MV, Gutiérrez OA. Rendimiento y calidades de chapa en clones de chopo a diferentes alturas del fuste. Investigación Agraria: Sistemas y Recursos Forestales 2002;11(2):311–23.
[35] Balatinecz JJ, Kretschmann DE, Leclercq A. Achievements in the utilization of poplarwood-guideposts for the future. Forestry Chronicle 2001;77(2):265–9. https://doi.org/10.5558/tfc77265-2.
[36] Basterra LA, Balmori JA, Morillas L, Acuña L, Casado M. Internal reinforcement of laminated duo beams of low-grade timber with GFRP sheets. Constr Build Mater 2017;154:914–20. https://doi.org/10.1016/j.conbuildmat.2017.08.007.
[37] Basterra LA, Acuña L, Casado M, Lopez G, Bueno A. Strength testing of Poplar duo beams, Populus x euramericana (Dode) Guinier cv. I-214, with fiber reinforcement. Constr Build Mater 2012;36:90–6. https://doi.org/10.1016/j.conbuildmat.2012.05.001.
[38] Garnica J. La importancia del chopo en la industria. Jornadas de Salicáceas-V Congreso Internacional de Salicáceas 2017:22.
[39] EN-1310. Round and sawn timber. Method of measurement of features; 1997. p. 26.
[40] EN-408. Timber structures. Structural timber and glued laminated timber-Determination of some physical and mechanical properties; 2011+A1:2012. p. 38.
[41] Wilcox R. Introduction to robust estimation and hypothesis testing. In: Introduction to Robust Estimation and Hypothesis Testing. Elsevier; 2017. p. 1–23. https://doi.org/10.1016/B978-0-12-804733-0.00001-9.
[42] Welch BL. On the comparison of several mean values: an alternative approach. Biometrika 1951;38(3-4):330–6. https://doi.org/10.1093/biomet/38.3-4.330.
[43] Du Mond C, Lenth R. A Robust Confidence Interval for Location. Technometrics 1987;29(2):211–9. https://doi.org/10.2307/1269776.
[44] R Development Core Team. The R Project for Statistical Computing. [En línea]; 2019. https://www.R-project.org/.
[45] Kuhn M, Johnson K. Applied Predictive Modeling. New York: Springer Science+Business Media; 2013. https://doi.org/10.1007/978-1-4614-6849-3.
[46] Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Softw 2008;28(5):1–26. https://doi.org/10.18637/jss.v028.i05.
[47] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97. https://doi.org/10.1007/BF00994018.
[48] Grenander U, Duda RO, Hart PE. Pattern classification and scene analysis. J Am Stat Assoc 1974:829. https://doi.org/10.2307/2286028.
[49] Friedman N, Geiger D, Goldszmit M. Bayesian Network Classifiers. Mach Learn 1997;29(2–3):131–63. https://doi.org/10.1023/A:1007465528199.
[50] Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Trans Inform Theory 1967;13(1):21–7. https://doi.org/10.1109/tit.1967.1053964.
[51] Pang SL, Gong JZ. C5.0 Classification algorithm and application on individual credit evaluation of banks. Systems Eng Theory-Practice SETP 2009;29(12): 94–104. https://doi.org/10.1016/S1874-8651(10)60092-0.
[52] Quinlan, JR. C5. [En línea]; 2007. [Citado el: 08 de April de 2019.] http://rulequest.com.
[53] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997;55(1):119–39. https://doi.org/10.1006/jcss.1997.1504.
[54] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943;5(4):115–33. https://doi.org/10.1007/BF02478259.
[55] Breiman L. Random Forests. Mach Learn 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324.
[56] Breiman L. Bagging predictors. Mach Learn 1996;24(2):123–40. https://doi.org/10.1007/BF00058655.
[57] Hastie T, Tibshirani R, Buja A. Flexible discriminant analysis by optimal scoring. J Am Stat Assoc 1994;89(428):1255–70. https://doi.org/10.1080/01621459.1994.10476866.
[58] Cox DR. Classification. Ghatak A. Machine Learning with R. Singapore: Springer Nature; 2017.
[59] Valiant LG. A theory of the learnable. Commun ACM 1984;27(11):1134–42.
[60] Breiman L. Prediction games and arcing algorithms. Neural Comput 1999;11(7): 1493–517. https://doi.org/10.1162/089976699300016106.
[61] Bro R, Smilde AK. Centering and scaling in component analysis. J Chemometr 2003;17(1):16–33. https://doi.org/10.1002/cem.773.
[62] Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. [En línea]; 2003 [Citado el: 20 de September de 2019] http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.
[63] Cerioli A, Riani M, Atkinson AC. Robust classification with categorical variables. In: Rizzi A, Vichi M, editors. Proceeding in Computational Statistics. Heidelberg: Physica-Verlag HD; 2006.
[64] D'yakonov AG. Solution methods for classification problems with categorical attributes. Appl Math 2014;46:81–8.
[65] Van Aken D, Pavlo A, Gordon GJ, Zhang B. Automatic Database Management System Tuning Through Large-scale Machine Learning. In: [SIGMOD '17-Proceedings of the 2017 ACM International Conference on Management of Data]; 2017. https://doi.org/10.1145/3035918.306.
[66] Larson S. The shrinkage of the coefficient of multiple correlation. J Educat Psychol 1931;22:45–55. https://doi.org/10.1037/h0072400.
[67] Branco P, Torgo L, Ribeiro R. Relevance-Based Evaluation Metrics for Multi-class Imbalanced Domains. [PAKDD 2017. Lecture Notes in Computer Science]; 2017. https://doi.org/10.1007/978-3-319-57454-7_54.
[68] Tharwat A. Classification assessment methods. Appl Comput Inform 2021;17(1): 168–92. https://doi.org/10.1016/j.aci.2018.08.003.
[69] Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The Balanced Accuracy and Its Posterior Distribution. [20th International Conference on Pattern Recognition]. Istanbul: s.n.; 2010. doi: 10.1109/ICPR.2010.764.
[70] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74. https://doi.org/10.2307/2529310.
[71] Gorodkin J. Comparing two k-category assignments by a k-category correlation coefficient. Comp Biol Chem 2004;28(5):367–74. https://doi.org/10.1016/j.compbiolchem.2004.09.006.
[72] Matthews BW. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. BBA-Protein Struct 1975;405(2):442–51. https://doi.org/10.1016/0005-2795(75)90109-9.
[73] Wei JM, Yuan XJ, Hu QH, Wang SQ. A novel measure for evaluating classifiers. Expert Syst Appl 2010;37(5):3799–809. https://doi.org/10.1016/j.eswa.2009.11.040.
[74] Delgado R, Núñez-González JD. Enhancing Confusion Entropy as Measure for Evaluating Classifiers. In: Graña M. et al., editors. [International Joint Conference SOCO'18-CISIS'18-ICEUTE'18. Advances in Intelligent Systems and Computing]; 2019. p. 771.
[75] Ballabio D, Grisoni F, Todeschini R. Multivariate comparison of classification performance measures. Chemometrics Intell Lab Syst 2018;174:33–44. https://doi.org/10.1016/j.chemolab.2017.12.004.
[76] Roblot G, Coudegnat D, Bleron L, Collet R. Evaluation of the visual stress grading standard on French Spruce (Picea excelsa) and Douglas-fir (Pseudotsuga menziesii) sawn timber. Ann For Sci 2008;65(8):4. https://doi.org/10.1051/forest:2008071.
[77] Llana DF, Arriaga F, Esteban M, Íñiguez-González G. Comparison between wet and dry timber visual strength grading according to the Spanish (UNE 56544) and German (DIN 4074–1) standards. Mater Construcc 2019;69(336):13. https://doi.org/10.3989/mc.2019.03319.
[78] Moya L, Cardoso A, Cagno M, O'Neill H. Structural characterization of pine lumber from Uruguay. Maderas, Cienc tecnol 2015;17(3):597–612. https://doi.org/10.4067/S0718-221X2015005000053.
[79] Acuña L, Spavento E, Casado MM, Basterra LA, Keil GD. Metodología de ensayo no destructivo aplicada a madera estructural de *Populus* x *euramericana* I-214 [Non-destructive testing techniques in *Populus* x *euramericana* I-214 structural timber]. In: Jornadas Sudamericanas de Ingeniería Estructural]. Montevideo, Uruguay: s.n; 2014. p. 12.
[80] Vega A, Dieste A, Guaita M, Majada J, Baño V. Modelling of the mechanical properties of Castanea sativa Mill. structural timber by a combination of non-

destructive variables and visual grading parameters. Eur J WoodModellierung der mechanischen Eigenschaften von Castanea sativa Mill. Bauholz durch eine Kombination von visuell und zerstörungsfrei bestimmten Parametern. Prod 2012; 70(6):839–44. https://doi.org/10.1007/s00107-012-0626-7.

[81] Casado Sanz M, Colunga Villar A, Acuña Rello L, Basterra Otero A, Ramón Cueto G, López Rodrigez G, et al. Clasificación de madera estructural de *Populus* x *euramericana* mediante ultrasonidos [Grading of structural timber of *Populus* x *euramericana* clone I-214 with ultrasonic]. [Proceedings of the 5°Congreso Forestal Español]. Ávila, Spain: s.n.; 2009 (a).

[82] Casado Sanz M, Escudero Diez I, Acuña Rello L, Vecilla García D, Basterra Otero LA, Ramón Cueto G, et al. Técnicas vibratorias aplicadas a madera estructural de *Populus* x *euramericana* [Application of vibration techniques to structural timber of *Populus* x *euramericana*]. [Proceedings of the 5°Congreso Forestal Español]. Ávila, Spain: s.n.; 2009 (b).

[83] Adell Almazán FJ, Hermoso Prieto E, ArriagaMartitegui F, Richter C. Comparison of the Spanish visual strength grading standard for structural sawn timber (UNE 56544) with the German one (DIN4074) for Scots pine (*Pinus sylvestris* L.) from Germany. Holz Roh Werkst 2008;66:253–8. https://doi.org/10.1007/s00107-008-0241-9.

[84] Conde M, Fernández-Golfín JI, Hermoso E. Improving the prediction of strength and rigidity of structural timber by combining ultrasound techniques with visual grading parameters. Mater Constr 2007;57(288):49–59. https://doi.org/10.3989/mc.2007.v57.i288.64.

[85] Arriaga Martitegui E, Esteban Herrero M, Arguelles Álvarez R, Bobadilla Maldonado I, Iñiguez GG. The effect of wanes on the bending strength of solid timber beams. Mater Constr 2007;57(288):61–76. https://doi.org/10.3989/mc.2007.v57.i288.65.

[86] Díez MR, Fernández-Golfín JI. Influencia de diversos factores en la calidad de la madera de uso estructural de *P. sylvestris* L. Invest. [Influence of various factors on *P. sylvestris* L. structural timber grade]. Forest System 1998;7(1–2):41–51. https://doi.org/10.5424/589.

[87] Donaldson L. Microfibril angle: measurement, variation and relationships: a review. IAWA J 2008;29(4):345–86. https://doi.org/10.1163/22941932-90000192.