

# Modelos de Regresión Logística Multinomial Ordinal y Redes Neuronales Artificiales para la clasificación de madera aserrada

## Multinomial and Ordinal Logistic Regression Models and Artificial Neural Networks for lumber grading

Ouorou Ganni Mariel Guera<sup>1</sup> • José Antônio Aleixo da Silva<sup>2</sup> • Rinaldo Luiz Caraciolo Ferreira<sup>2</sup>  
• Daniel Álvarez Lazo<sup>3</sup> • Héctor Barrero Medel<sup>3</sup> • Madelén C. Garofalo Novo<sup>4</sup>  
• Moacyr Cunha Filho<sup>1</sup> • José Wesley Lima Silva<sup>1</sup>

Recibido: 3/8/2020

Aceptado: 29/1/2021

Publicado: 29/7/2021

### Abstract

Lumber classification is one of the most subjective activities of the final phase of log sawing process in sawmills. The objective of this research was to propose tools that assist in conifers lumber grading. The research was carried out at the sawmill Combate de Tenerías of Macurije integrated forest company, Pinar del Río, Cuba. The data used comes from 259 lumber pieces of *Pinus caribaea* var. *caribaea* classified following the requirements (24 variables) and classes established by the conifers lumber grader used in Cuba. We proceeded to fit a Multinomial and Ordinal Logistic Regression (MOLR) model and train Artificial Neural Networks (ANNs). The parameters of the MOLR model were estimated using the maximum likelihood method optimized with the Newton-Raphson algorithm. ANNs were trained with Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. The multicollinearity problem, usually present in modeling with numerous predictive variables, was addressed with factor analysis, using the factors retained as inputs to the models. Based on the percentage of correct classification, the ANN RBF 24-8-4 was superior to the ordinal logistic regression equations.

**Key words:** Factorial Analysis, Radial Basis Function, Multilayer Perceptron.

1. Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco (PPGBEA/UFRPE); Recife-PE, Brasil; [gueraforest@gmail.com](mailto:gueraforest@gmail.com), [moacyr.cunhafo@ufrpe.br](mailto:moacyr.cunhafo@ufrpe.br), [wesleylima\\_16silva@hotmail.com](mailto:wesleylima_16silva@hotmail.com)

2. Departamento de Ciência Florestal, Laboratório de Biometria e Manejo Florestal (LBMF), Universidade Federal Rural de Pernambuco (DCFL/UFRPE); Recife-PE, Brasil; [jaaleixo@gmail.com](mailto:jaaleixo@gmail.com), [rinaldo.ferreira@ufrpe.br](mailto:rinaldo.ferreira@ufrpe.br)

3. Departamento Forestal, Universidad de Pinar del Río “Hermanos Saíz Montes de Oca”, Pinar del Río, Cuba; [daniel@upr.edu.cu](mailto:daniel@upr.edu.cu), [hbarrero@upr.edu.cu](mailto:hbarrero@upr.edu.cu)

4. Departamento de Matemáticas, Universidad de Pinar del Río “Hermanos Saíz Montes de Oca”, Pinar del Río, Cuba; [madelen@upr.cuedu.cu](mailto:madelen@upr.cuedu.cu)

## Resumen

La clasificación de la madera aserrada es una de las actividades más subjetivas de la fase final del proceso de aserrado de trozas en los aserríos. La presente investigación tuvo como objetivo proponer herramientas de auxilio en la clasificación de madera aserrada de coníferas. El estudio se realizó en el aserrío Combate de Tenerías de la Empresa Forestal Integral (EFI) Macurije, Pinar del Río, Cuba. Los datos utilizados proceden de 259 tablas de *Pinus caribaea* var. *caribaea* clasificadas siguiendo las exigencias (24 variables) y clases establecidas por el clasificador de madera aserrada de coníferas utilizado en Cuba. Se procedió al ajuste de un modelo de Regresión Logística Multinomial Ordinal (RLMO) y al entrenamiento de Redes Neuronales Artificiales (RNAs). Los parámetros del modelo RLMO fueron estimados usando el método de máxima verosimilitud optimizada con el algoritmo de Newton-Raphson. El entrenamiento de las RNAs se realizó con el algoritmo Broyden-Fletcher-Goldfarb-Shanno (BFGS). El problema de multicolinealidad, casi siempre presente en modelaciones con numerosas variables predictivas, fue abordado con el análisis factorial, utilizando los factores retenidos como entradas de los modelos. Con base en el porcentaje de aciertos en la clasificación, la RNA RBF 24-8-4 fue superior a las ecuaciones de regresión logística ordinal.

**Key words:** Análisis Factorial, Función de Base Radial, Perceptrón Multicapa.

## Introducción

La madera, uno de los pocos recursos naturales renovables, es predominante en nuestra vida cotidiana, así como en los procesos económicos [9]. Una transformación mecánica primaria que este recurso comúnmente sufre para adquirir mayor utilidad y mayor valor agregado es el proceso de aserrado que ocurre generalmente en aserraderos.

Las propiedades físicas y mecánicas de este recurso natural pueden variar de forma inter e intraespecífica y en función de su posición en el mismo árbol. Estas propiedades intrínsecas, sumadas a factores ambientales y variaciones oriundas del aserrado y secado, proporcionan maderas aserradas de diversas calidades, generando la necesidad de su clasificación por clase de calidad para un uso más adecuado. Sin embargo, irregularidades, patrones no repetidos y defectos dificultan esa clasificación [18].

La clasificación es generalmente realizada por reglas o normas estandarizadas en todo el territorio nacional de un país [17]. Hay dos tipos de clasificación: la visual y la mecánica [12]. La calificación visual ocurre casi siempre en el aserradero [17] y se realiza de acuerdo con una determinada norma de clasificación generalmente basada en el número, tamaño y localización de defectos presentes en las piezas [2]. Según [5], esta clasificación se basa en la premisa de que los defectos pueden ser observados y evaluados a simple vista. Su eficiencia depende de la experiencia del clasificador. La clasificación mecánica se realiza con pruebas no destructivas acompañadas de un análisis visual. Si bien este último tipo de clasificación es más eficiente, según [4], es costoso y demorado, razón por la cual la primera es ampliamente más usada. Por lo tanto, una clasificación visual eficiente y menos subjetiva es requerida.

Numerosas normas de clasificación de madera aserrada fueron propuestas e implementadas por países, grupo de especies (coníferas y latifolias) e incluso por especie. Una de las más conocidas es la norma americana, ASTM D245 (Standard Practice for Establishing Structural Grades and Related Allowable Properties for Visually Graded Lumber), propuesta específicamente para una clasificación visual [6]. En Cuba, una de las normas utilizadas para clasificar la madera aserrada de coníferas es el clasificador propuesto por [11], el cual se basa en 24 variables. La evaluación visual de las exigencias técnicas contempladas en las normas y la decisión de la inclusión o no de una pieza en una determinada clase de calidad puede ser subjetiva, principalmente para piezas con características que la posicionan en el límite entre dos clases.

La alta velocidad (hasta 5 m.s<sup>-1</sup>) a la que las trozas son generalmente procesadas en los aserraderos exige una clasificación rápida y eficaz de la madera aserrada, la cual no es posible sin automatización de procesos [15]. Los aserríos cubanos, incluyendo al aserrío Combate de Tenerías, objeto del presente estudio, son semi-automatizados. El proceso de aserrado inicia con una entrada automática de madera (trozas) y finaliza con la clasificación manual de la madera aserrada. La lentitud de esta clasificación manual confrontada a esta velocidad de aserrado obliga a los clasificadores a trabajar bajo presión, lo que aumenta la probabilidad de error en la clasificación. Esta clasificación se puede automatizar mediante la utilización de un medio informático rápido y preciso de clasificación de la madera aserrada [21]. Estos sistemas automatizados anunciados por [15] ya son funcionales en numerosos aserraderos. Los mismos incluyen complejos mecanismos compuestos por cámaras, luces, láseres, rayos X, computadoras y otros dispositivos necesarios para identificar defectos en las piezas de madera aserrada. Para la automatización de este proceso en el aserrío Combate de Tenerías,

un sistema similar al desarrollado por [24] puede ser montado y entrenado para el reconocimiento de los defectos en madera aserrada. El conjunto de salida de este sistema alimentará modelos cuyas entradas son los defectos identificados por un sistema de barrido láser. Según [16], esta automatización es imprescindible para agilizar el proceso y reducir la tasa de reclasificación de la madera que surge de divergencias recurrentes entre vendedores (los aserraderos) y compradores exigentes.

Entre las técnicas utilizadas para auxiliar el proceso de clasificación, se destacan: el modelo de regresión logística multinomial ordinal, cuya utilización ha crecido en los últimos años en el análisis de datos y cuya respuesta es presentada de forma categórica y ordenada [1]; y las Redes Neuronales Artificiales (RNAs) cuya estructura permite la clasificación de datos en categorías o clases [7].

Ante el evidente problema de subjetividad en la clasificación visual de madera aserrada arriba expuesto, el presente trabajo tuvo como objetivo, obtener herramientas de auxilio a la clasificación de madera aserrada de *Pinus caribaea* var. *caribaea* en el aserrío Combate de Tenerías, por medio de ajustes de modelos de Regresión Logística Multinomial Ordinal y configuración de RNAs.

## Materiales y métodos

### Caracterización del local de estudio

El presente estudio se realizó en el aserrío Combate de Tenerías perteneciente a la Empresa Forestal Integral

**Cuadro 1.** Predictores de clases de calidad de madera aserrada de coníferas.

**Table 1.** Predictors of conifers sawnwood quality classes.

Número	Variables	Símbolos
1	Diámetros de nudos sanos y adheridos (cm)	DNSA
2	Nudos adheridos parcialmente y no adheridos (cantidad)	NSPAnA
3	Nudos sueltos, parcialmente podridos y podridos (cantidad)	NSPPeP
4	Diámetro de las grietas (cm)	DRch
5	Ancho de las grietas (cm)	LRch
6	Desviación de las fibras (%)	DFibr
7	Medula simple y doble (cantidad)	MSeD
8	Bolsas de resina (cantidad / metro lineal)	BR
9	Ancho de las bolsas de casca (cm)	LBC
10	Longitud de las bolsas de casca (cm)	CBC
11	Podredumbre externa reducida a polvo (cantidad)	PERP
12	Manchas de hongos en el Cerne (%)	MFC
13	Orificios causados por insectos (cantidad / metro lineal)	OCI
14	Daños mecánicos (cantidad)	DM
15	Gema (%)	G
16	Curvatura del ancho (%)	CC
17	Curvatura del largo (%)	CL
18	Curvatura en la longitud (%)	Ccomp
19	Desviaciones positivas de aserrado permisibles en la longitud (DPSM) (cm)	DPSM_C+
20	Desviaciones negativas de aserrado permisibles en la longitud (DPSM) (cm)	DPSM_C-
21	Desviaciones positivas de aserrado permisibles en el ancho (DPSM) (cm)	DPSM_L+
22	Desviaciones negativas de aserrado permisibles en el ancho (DPSM) (cm)	DPSM_L-
23	Desviaciones positivas de aserrado permisibles en el grosor (DPSM) (cm)	DPSM_G+
24	Desviaciones negativas de aserrado permisibles em el grosor (DPSM) (cm)	DPSM_G-

Fuente: Clasificador de madera aserrada de coníferas de [11]

(EFI) Macurije ubicada en Pinar del Río, Cuba. La especie objeto de estudio fue el *Pinus caribaea* var. *caribaea* que representa el 90 % de la madera procesada en este aserrío [13]. El predominio de esta especie en el sistema de producción de la empresa está en consonancia con las estadísticas nacionales según las cuales esta ocupa el primer lugar en los planes de reforestación y en los sistemas de producción de la isla [19]. La tecnología instalada en el aserradero, con sierras de diferentes calibres y anchos, permite utilizar muchos esquemas de aserrado en dos líneas de producción: una de grandes dimensiones (trozas de 4 m) y otra de pequeñas dimensiones (trozas de 3 m) que operan con sierras de cintas verticales, dobles y de madera móvil durante 8 horas diarias.

### Recolección de datos

Los datos utilizados en el estudio se obtuvieron durante el aserrado de trozas de *Pinus caribaea* var. *caribaea* en el aserradero Combate de Tenerías en el mes de marzo de 2015. Estos datos proceden de la minuciosa clasificación visual de 259 tablas obtenidas al final del proceso de aserrado. Las variables independientes o predictoras consideradas en la elaboración de los modelos de predicción de la calidad de madera aserrada de *Pinus caribaea* var. *caribaea*, tanto los de regresión como los de RNAs, son las exigencias técnicas (variables) del clasificador de madera aserrada de coníferas de [11] (Cuadro 1). La variable dependiente fue la calidad, una variable categórica que cuenta con cuatro categorías. En orden decreciente, esas categorías son: selecta, Clase I, Clase II y Clase III. La clasificación de una pieza en determinada clase de calidad depende de la cantidad de defectos registrados en la misma. La probabilidad de que una pieza sea clasificada en la clase SELECTA disminuye con el aumento de la cantidad de defectos en esta. Los límites de inclusión de las piezas en cada clase de calidad se encuentran en el clasificador de [11].

### Modelo de Regresión Logística Multinomial Ordinal (RLMO)

Dado que la variable respuesta (calidad) de la clasificación de la madera aserrada es discreta, con más de dos categorías ordenadas de mejor a peor (Selecta, Clase I, Clase II, Clase III), el modelo de regresión más adecuado para la clasificación es el de regresión logística ordinal también conocido como modelo de regresión logit-ordinal. Este modelo tiene una parametrización diferente del modelo logístico ordinario, ya sea binario o multinomial. La modelación se centra en la probabilidad de elegir una de las J categorías conociendo algunas variables explicativas y se basa en las probabilidades acumulativas. En el modelo RLMO, la probabilidad de que una pieza pertenezca a una categoría j o a una

categoría inferior, ya que la misma es acumulativa, se determinó por medio de las ecuaciones 1 y 2.

$$\text{Log} \left[ \frac{p(y \leq j|x_i)}{p(y > j|x_i)} \right] = \alpha_j + \beta X_i + \varepsilon_i \quad (1)$$

$$p(y \leq j|x_i) = \frac{\text{Exp}(\alpha_j + \beta X_i)}{1 + \text{Exp}(\alpha_j + \beta X_i)} + \varepsilon \quad (2)$$

Donde: y es la variable dependiente con j categorías;  $x_i$  es la i-ésima variable predictor;  $\alpha_j$  es el parámetro de la j-ésima categoría y  $\beta$  es el parámetro común a todas las categorías.

La ecuación de RLMO fue estimada usando el método de máxima verosimilitud (Ecuación 3) optimizada con el algoritmo de Newton-Raphson en XLSTAT versión 2016.

$$l(\alpha, \beta) = \sum_{i=1}^n \sum_{j=1}^J y_{ij} \text{Log}[p(y \leq j|x_i) - p(y \leq j-1|x_i)] \quad (3)$$

Donde: y es la variable dependiente con j categorías;  $x_i$  es la i-ésima variable predictor;  $\alpha_j$  es el parámetro de la j-ésima categoría y  $\beta$  es un parámetro común a las categorías.

La significancia del modelo independiente (sólo el intercepto) fue analizada por la prueba de la Razón de Verosimilitudes (TRV) (Ecuaciones 4 y 5). En el mismo, se probó también si todos los coeficientes de regresión son todos nulos con excepción de la interceptación ( $b_0$ ). En caso de rechazar la hipótesis nula, se concluye que al menos uno de los coeficientes es estadísticamente distinto de cero. La significancia de esta prueba conduce a la realización de la prueba de Wald para cada coeficiente (Ecuación 6).

$$D = -2 \text{Ln} \left[ \frac{\text{FMVMC}}{\text{FMVMS}} \right] \quad (4)$$

$$D = -2 \sum_{i=1}^n \left[ y_i \text{Ln} \left( \frac{\pi_i}{y_i} \right) + (1 - y_i) \text{Ln} \left( \frac{1 - \pi_i}{1 - y_i} \right) \right] \quad (5)$$

$$W_j = \frac{b_j}{\text{Var}(b_j)} \quad (6)$$

Donde: FMVMC= Función de Máxima Verosimilitud del Modelo Corriente (sólo variables deseadas); FMVMS= Función de Máxima Verosimilitud del Modelo Saturado

(incluye todas las variables);  $b_j$  es la estimación del parámetro asociado a la categoría  $j$ ;  $Var(b_j)$  es la varianza de la estimación del parámetro asociado a la categoría  $j$ .

**Análisis factorial con las variables predictoras de la calidad de madera aserrada**

Dada la gran cantidad de variables que involucra la clasificación de madera aserrada, la probabilidad de que exista multicolinealidad entre los predictores del modelo es alta. Aunque la presencia de este fenómeno genere ajustes aparentemente buenos, con altos coeficientes de determinación ( $R^2$ ), esta afecta significativamente la capacidad predictiva de los modelos elaborados. Para evitarla, se realizó un análisis factorial para reducir la dimensión del conjunto de datos sin una significativa pérdida de información. Los factores obtenidos se utilizaron como entradas de los modelos logit-ordinal.

El análisis factorial siguió los siguientes pasos: (1) suficiencia muestral determinada por [14], según el cual, la razón entre el número de observaciones y la cantidad de variables debe exceder de cinco a uno o más; (2) verificación de la adecuación de la aplicación de la técnica por la prueba de KMO (Kaiser-Meyer-Olkin), cuya interpretación se hizo con base en [10]. Según este autor, un KMO entre 0,90 y 1 es considerado excelente; entre 0,80 y 0,89 es bueno; entre 0,70 y 0,79 es mediano; entre 0,60 y 0,69 es mediocre; entre 0,50 y 0,59 es malo y entre 0 y 0,49 es totalmente inadecuado. La técnica de extracción de factores utilizada fue la de componentes principales. Buscando un mayor grado de parsimonia, una mayor varianza explicada y cumpliendo con la exigencia de [20] de tener entre 20 y 50 variables, el número óptimo de factores fue determinado por el criterio de Kaiser, el cual establece que todos los factores extraídos tienen que presentar un autovalor por encima de 1 (uno). Para una mejor interpretación de los factores, se realizó una rotación ortogonal varimax, método propuesto por Kaiser [8]. El análisis factorial se realizó con el Sistema de Análisis Estadístico – SAS (Versión 9.4, SAS Institute, NC).

**Redes Neuronales artificiales para la clasificación de madera aserrada**

Los entrenamientos de las RNAs se realizaron con el algoritmo Broyden-Fletcher-Goldfarb-Shanno (BFGS) conocido por su mejor tasa de convergencia y su capacidad de búsqueda más inteligente [3]. Las RNAs entrenadas fueron del tipo Radial Basis Function (RBF) y Multi-Layer Perceptron (MLP) y las funciones de activación para las neuronas ocultas y de salida fueron las funciones seno, exponencial, identidad, logística y tangente hiperbólica. La base de datos se dividió en tres subconjuntos independientes con las siguientes

proporciones: 50 % para el entrenamiento, 25 % para la validación y 25 % para la prueba. Para facilitar la convergencia de las RNAs, las variables se normalizaron para los intervalos [0,1] o [-1,1] a través de la transformación lineal. Los entrenamientos se realizaron en el programa Statistica versión 8.0 (Stat Soft, Inc., Tulsa, OK, USA).

El análisis de la importancia de cada una de las variables predictoras en la clasificación de madera aserrada, también conocido como análisis de sensibilidad, se realizó calculando la contribución de cada variable  $j$  con las ecuaciones 7 y 8 [22].

$$Cont_j(\%) = \frac{\sum_{n=1}^N Sen_{jn}^2}{\sum_{j=1}^{N_{inp}} \sum_{n=1}^N Sen_{jn}^2} \tag{7}$$

$$Sen_{jn} = \sum_{k=1}^{N_{hid}} W_{kj} \cdot f'(net_k) \cdot \sum_{i=1}^{N_{out}} W_{ik} \cdot f'(net_i) \cdot e_i(n) \tag{8}$$

En que:  $Cont_j(\%)$ = contribución de cada variable;  $Sen_{jn}$  = sensibilidad para cada variable de entrada  $j$  ( $j=1, \dots, N_{inp}$ ) con relación a la salida para un dado ejemplo  $n$  ( $1, \dots, N$ );  $N_{inp}$  = número total de neuronas o variables en la capa de entrada;  $N_{out}$  = número de neuronas en la capa de salida.  $W_{kj}$  = pesos sinápticos que conectan la capa oculta a la capa de entrada y  $W_{ik}$  = pesos sinápticos que conectan la capa de salida a la capa oculta.  $f'(net_j)$  y  $f'(net_k)$  son las derivadas de las funciones de activación de las neuronas de las capas de salida y oculta, respectivamente;  $e_i(n)$ = error en la capa de salida.

**Criterios de evaluación y selección de los modelos de regresión y de las RNAs**

Las mejores RNAs fueron las que presentaron un mayor porcentaje de acierto de clasificación. Los modelos de regresión logística polinómica ordinal, basándose en estimaciones de la máxima verosimilitud, los criterios utilizados para su evaluación fueron: el criterio de información bayesiano - BIC o SBC (Ecuación 9) y los pseudo- $R^2$  como el  $R^2$ -McFadden (Ecuación 10); el  $R^2$ -Cox y Snell (Ecuación 11) y el  $R^2$  - Nagelkerke (Ecuación 12). Estos pseudo- $R^2$  se interpretan similarmente al coeficiente de determinación ( $R^2$ ) de un análisis de regresión ordinaria. Siendo así, análogamente, cuando mayores sean sus valores, mejor es el ajuste.

$$BIC = -2 \log f(X_n \parallel \theta) + p \log n \tag{9}$$

En que:  $f(x_n \parallel \theta)$  es el modelo elegido,  $p$  y  $n$  son el número de parámetros y de observaciones, respectivamente.

**Cuadro 2.** Estadísticas del ajuste del modelo de regresión logística politómica ordinal.

**Table 2.** Fit statistics of the ordinal and polytomous Logistic Regression Model.

Estadística	Independiente	Completo
-2 Log (Verosimilitud)	648,955	100,025
R <sup>2</sup> (McFadden)	0,000	0,322
R <sup>2</sup> (Cox e Snell)	0,000	0,311
R <sup>2</sup> (Nagelkerke)	0,000	0,323
BIC	665,626	383,446

**Cuadro 3.** Prueba de la significancia de los parámetros del modelo Logit ordinal.

**Table 3.** The ordinal logit model parameters significance test.

Estadística	Chi cuadrado	Pr > Chi <sup>2</sup>
-2 Log (Verosimilitud)	648,930	< 0,0001
Score	692,166	< 0,0001
Wald	0,952	1,000

$$R_{MF}^2 = 1 - \frac{L_f}{L_0} \tag{10}$$

$$R_{CS}^2 = 1 - \exp \left[ -\frac{2}{n} (L_f - L_0) \right] \tag{11}$$

$$R_N^2 = \frac{R_{CS}^2}{R_{MAX}^2} \quad \text{con} \quad R_{MAX}^2 = 1 - \exp \left[ \frac{2}{n} (L_0) \right] \tag{12}$$

Donde:  $L_0$  = Log-verosimilitud del modelo nulo;  $L_f$  = Log-verosimilitud del modelo final.

de que al menos un coeficiente, diferente del intercepto, no es nulo. Esta significancia condujo a la realización de la prueba de Wald, el cual fue significativo sólo para la variable DPSM\_C+ (desviaciones positivas permisibles en la longitud de la madera). Esto indica que en la ecuación logit-ordinal obtenida, sólo esta variable tiene una relación estadísticamente significativa con la calidad de la madera aserrada. Se puede concluir entonces una falta de consistencia de las estimaciones de los parámetros. En cuanto a los pseudo-R<sup>2</sup> (Cuadro 2), sus valores indican que un buen porcentaje de varianza de la calidad de la madera aserrada es explicado por las ecuaciones de regresión logística ordinal obtenidas. Los valores de BIC indican que el modelo completo (intercepto + variables predictoras) es superior al modelo independiente (sólo el intercepto).

## Resultados y discusión

### Modelo de regresión logística ordinal para la clasificación de madera aserrada de *Pinus caribaea* var. *caribaea*

Un previo análisis de multicolinealidad indicó la necesidad de excluir dos variables de la base de datos: las desviaciones positivas permisibles en el ancho (DPSM\_L+) y las desviaciones positivas permisibles en el grosor (DPSM\_G+). La significancia de la prueba de la razón de verosimilitudes (Cuadro 3) es un indicativo

La validación de la ecuación de regresión logística ordinal obtenida indicó un porcentaje de clasificación correcta del 54,44 % (Cuadro 4). Se supone que este porcentaje de acierto relativamente bajo es causado por la no significancia de la participación de la mayoría de las variables en el modelo.

Para abordar el problema de no significancia de la casi totalidad de los parámetros del modelo, la introducción de los datos de una forma más reducida fue considerada. Por ello, se realizó un análisis factorial descrito en la sección subsecuente para la obtención de factores cuyos scores serían las nuevas entradas del modelo.

**Cuadro 4.** Validación del modelo de regresión logística ordinal usando las 22 variables.

**Table 4.** Validation of the ordinal logistic regression model using the 22 variables.

De \ a	CLASE I	CLASE II	CLASE III	SELECTA	Total	% correcto
CLASE I	64	27	0	25	116	55,17
CLASE II	9	23	17	0	49	46,94
CLASE III	0	20	43	6	69	62,32
SELECTA	6	0	8	11	25	44,00
Total	79	70	68	42	259	54,44

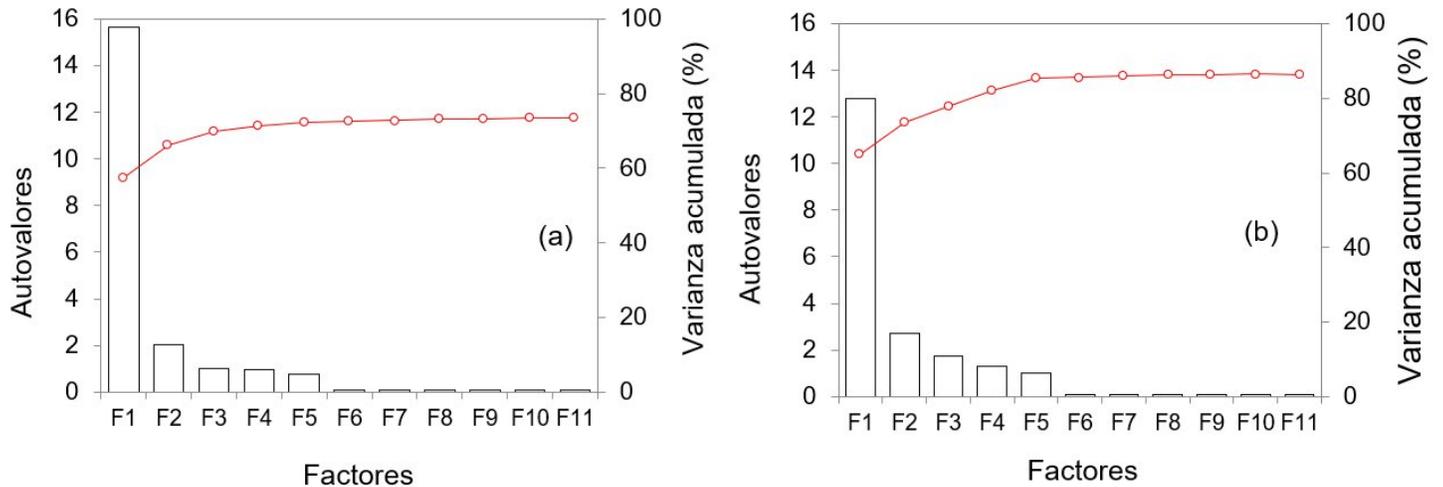


Figura 1. Gráfica de sedimentación de los autovalores antes (a) y después (b) de la rotación varimax.

Figure 1. Eigenvalues scree plot before (a) and after (b) the varimax rotation.

### Análisis factorial para la reducción de la dimensionalidad del conjunto de entrada

El valor del KMO de la muestra fue de 0,962, considerado excelente según la clasificación de [10]. Este valor indica que los datos son adecuados para la aplicación del análisis factorial. El criterio de Kaiser sugirió la extracción de tres factores. Estos tres factores explican el 77,93 % de la varianza total de las variables. La aplicación del método de rotación VARIMAX (Figura 1 y Cuadro 5) permitió la inclusión de otros dos factores que proporcionaron el aumento de la varianza total explicada del 77,93 % al 85,3 %.

El primer factor presentó un autovalor de 12,77, responsable por el 65,13 % de la varianza total; el segundo factor presentó un autovalor de 2,70 y explica cerca de 8,47 % de la varianza; el tercero, con un autovalor de 1,73, es responsable del 4,32 % de la varianza; el cuarto con un autovalor de 1,28 es responsable del 4,10 % de la varianza y el quinto y último, con un autovalor de 1,005, es responsable del 3,30 % de la varianza.

Se puede observar que después de la rotación varimax, el mayor cambio ocurrió al alrededor de los cinco primeros factores. La rotación permitió una redistribución de la varianza entre factores obteniendo así cinco factores con los autovalores encima de uno. Por ello, los cinco primeros factores se retuvieron para describir las relaciones de covarianza entre las variables involucradas.

El primer factor (F1) presentó una fuerte correlación positiva con las variables DPSM\_C+, DPSM\_C-, DPSM\_L+, DPSM\_L-, DPSM\_G+, DPSM\_G-, DPSM\_G+ y G (Cuadro 5). La naturaleza de estas variables permite

asociar ese factor con los defectos de aserrado en la madera aserrada (Cuadro 6). Las variables NSPANa, BR, PERP y DM se correlacionaron fuertemente con el Factor F2 asociado con los defectos de consistencia de la madera aserrada (Cuadros 5 y 6). El factor F3 está correlacionado con las variables LBC y CBC y por la naturaleza de esas variables, fue asociado a los defectos de estética de la madera aserrada (Cuadros 6 y 7). Las variables DRch y CC están fuertemente correlacionadas con el factor F4 y la variable CL con el factor F5, etiquetando estos dos factores como descriptores de los defectos de secado (Cuadros 5 y 6).

### Modelo de regresión logística ordinal basado en los factores

Los ajustes obtenidos con los factores fueron similares a los anteriores, con una diferencia en el BIC, el cual fue mejor con la introducción de los factores (Cuadros 7 y 8).

A diferencia del modelo anterior, en este modelo, todas las variables predictoras (factores seleccionados por el criterio de Kaiser) fueron significativos (Cuadro 9).

La contribución de esos factores siendo significativa, las estimaciones de los parámetros asociados (Cuadro 10) pueden ser consideradas confiables y su interpretación realizada.

En la clase II, para el factor F1 relacionado con los defectos de aserrado, para un aumento de una unidad se espera un aumento de 14,806 en las probabilidades de estar en una clase peor (clase III) cuando todos los otros factores se mantienen constantes. Esta

**Cuadro 5.** Matriz de cargas factoriales rotacionadas para las 24 variables predictoras.**Table 5.** Matrix of rotated factor loadings for the 24 predictors.

Variables	Factores rotacionados - Método Virimax					C
	F1	F2	F3	F4	F5	
DNSA	0,826	0,498	0,021	0,025	-0,013	0,932
NSPAnA	0,321	0,605	0,234	0,125	-0,004	0,540
NSPPeP	0,755	0,568	-0,049	-0,067	-0,002	0,899
DRch	-0,251	0,157	0,569	0,552	0,024	0,717
LRch	0,834	0,327	0,083	0,073	-0,035	0,816
DFibr	0,804	0,391	-0,018	-0,041	0,004	0,802
MSeD	0,623	0,407	0,278	0,285	0,021	0,713
BR	0,701	0,554	0,139	0,100	-0,038	0,829
LBC	0,157	0,093	0,754	0,019	-0,052	0,604
CBC	0,156	-0,058	0,732	0,082	0,136	0,588
PERP	0,561	0,740	-0,025	-0,012	-0,003	0,863
MFC	0,809	0,502	-0,066	-0,064	-0,009	0,916
OCI	0,828	0,503	0,054	0,015	-0,010	0,942
DM	0,552	0,748	-0,052	0,037	-0,003	0,868
G	0,833	0,321	0,088	0,032	-0,043	0,808
CC	0,226	0,011	0,090	0,895	0,027	0,861
CL	-0,036	-0,006	0,084	0,033	0,989	0,987
Ccomp	0,807	0,536	0,033	0,021	-0,002	0,940
DPSM_C+	0,951	0,221	0,133	0,085	-0,003	0,978
DPSM_C-	0,944	0,240	0,133	0,091	-0,004	0,975
DPSM_L+	0,949	0,212	0,139	0,087	-0,005	0,972
DPSM_L-	0,950	0,200	0,127	0,072	-0,006	0,965
DPSM_G+	0,950	0,215	0,121	0,071	-0,009	0,969
DPSM_G-	0,950	0,202	0,149	0,091	-0,005	0,974
Autovalores	15,633	2,033	1,037	0,974	0,783	-
Autovalores (Rotacionados)	12,766	2,700	1,731	1,281	1,005	-
Varianza (%)	65,1	8,5	4,3	4,1	3,3	-
Varianza acumulada (%)	65,1	73,6	77,9	82,0	85,3	-

F1, F2, F3, F4 y F5 son los factores extraídos y C es la comunalidad.

**Cuadro 6.** Nomenclatura de los factores y variables explicativas asociadas.**Table 6.** Nomenclature of factors and associated explanatory variables.

Factores	Variables asociadas	Nomenclatura
01	DPSM_C+; DPSM_C-; DPSM_L+; DPSM_L-; DPSM_G+; DPSM_G-, G	Defectos de aserrado
02	NSPAnA; NSPPeP; BR; PERP; DM	Defectos de consistencia
03	LBC; CBC	Defectos de estética
04	DRch; CC; CL	Defectos de secado
05		

**Cuadro 7.** Bondad de ajuste del modelo de regresión logística ordinal basada en factores.

**Table 7.** Goodness of fit of the ordinal logistic regression model based on factors.

Estadística	Independiente	Completo
-2 Log (Verosimilitud)	648,955	100,117
R <sup>2</sup> (McFadden)	0,000	0,688
R <sup>2</sup> (Cox e Snell)	0,000	0,602
R <sup>2</sup> (Nagelkerke)	0,000	0,689
BIC	665,626	100,140

**Cuadro 8.** Prueba de hipótesis para el modelo RLMO basada en factores.

**Table 8.** Hypothesis test for the RLMO model based on factors .

Estadística	GL	Chi-cuadrado	Pr > Chi <sup>2</sup>
-2 Log (Verosimilitud)	15	648,838	< 0,0001
Score	15	591,048	< 0,0001
Wald	15	2,627	1,000

**Cuadro 9.** Significación de los factores (análisis del tipo II).

**Table 9.** Significance of the factors (type II analysis).

Fuente	GL	Chi-cuadrado (Wald)	Pr > Wald	Chi-cuadrado (LR)	Pr > LR
F1	3	1,152	0,764	129,633	< 0,0001
F2	3	0,264	0,967	144,071	< 0,0001
F3	3	0,159	0,984	10,087	< 0,0001
F4	3	0,190	0,979	5,113	< 0,0001
F5	3	0,033	0,998	4,117	< 0,0001

**Cuadro 10.** Estimaciones de parámetros del modelo de RLMO basado en los factores.

**Table 10.** Parameter estimates of the RLMO model based on factors.

Categoría	Fuente	Valor	Desvío padrón	Chi-cuadrado de Wald	Pr > Chi <sup>2</sup>	Odds ratio
CLASE II	Intercepto	-2,200	24,485	0,008	0,928	
	F1	14,806	16,742	0,782	0,377	269,168.104
	F2	0,133	33,892	0,000	0,997	1,142
	F3	-5,453	20,941	0,068	0,795	0,004
	F4	-3,952	16,409	0,058	0,810	0,019
	F5	-2,763	32,898	0,007	0,933	0,063
CLASE III	Intercepto	-8,207	33,492	0,060	0,806	
	F1	24,854	23,857	1,085	0,298	622,158.108
	F2	10,486	39,142	0,072	0,789	35825,115
	F3	-3,253	29,683	0,012	0,913	0,039
	F4	-1,593	22,408	0,005	0,943	0,203
	F5	1,221	9,531	0,016	0,898	3,389
SELECTA	Intercepto	-6,602	28,066	0,055	0,814	
	F1	6,323	49,219	0,017	0,898	557,475
	F2	1,703	55,595	0,001	0,976	5,490
	F3	-8,500	27,694	0,094	0,759	0,000
	F4	-6,144	16,185	0,144	0,704	0,002
	F5	0,665	11,858	0,003	0,955	1,944

**Cuadro 11.** Validación del modelo RLMO basado en los factores.

**Table 11.** Validation of the RLMO model based on factors.

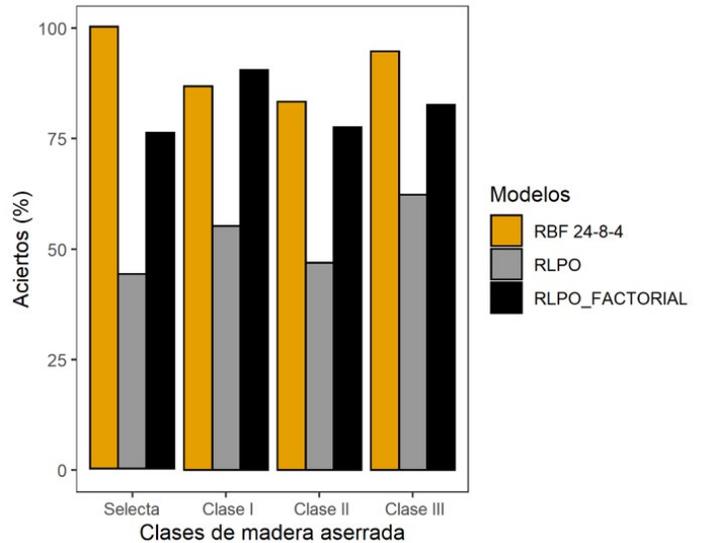
De \ a	CLASE I	CLASE II	CLASE III	SELECTA	Total	% correcto
CLASE I	105	11	0	0	116	90,52%
CLASE II	6	38	5	0	49	77,55%
CLASE III	2	10	57	0	69	82,61%
SELECTA	6	0	0	19	25	76,00%
Total	119	59	62	19	259	84,55%

interpretación se puede realizar en todas las clases para analizar la influencia de los factores en la inclusión o no de las piezas de madera aserrada en determinada clase. El aumento del porcentaje de acierto de 54,44% (Cuadro 4) a 84,55 % (Cuadro 11) indica consistencia y mejor capacidad predictiva del modelo ajustado con los factores.

**RNAs para la clasificación de la madera aserrada de *Pinus caribaea* var. *caribaea***

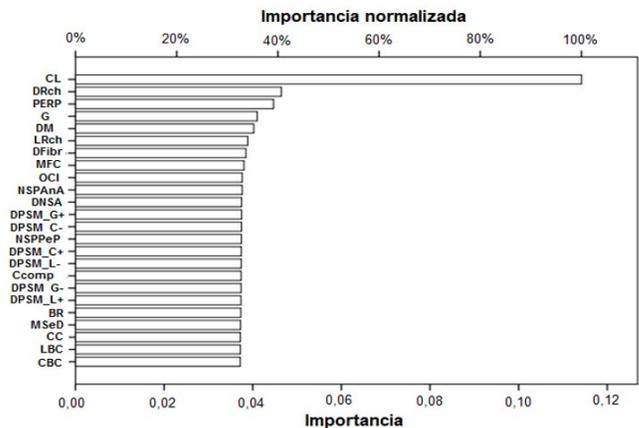
Las cinco mejores RNAs fueron del tipo RBF (Cuadro 12). Este resultado está en consonancia con [22] que alega que este tipo de RNAs (RBFs) son generalmente superiores a las del tipo MLP en las tareas de clasificación. Esta superioridad se atribuye a su propia estructura que requiere un entrenamiento no supervisado para determinar las estimaciones de los parámetros de las funciones de base radial de la capa oculta y un entrenamiento supervisado para ajustar los pesos que unen la capa oculta a la de salida. En el entrenamiento no supervisado, el algoritmo de clusterización utilizado permitió la identificación de grupos potenciales, con la capacidad de identificación de relaciones no lineales. La segunda fase del entrenamiento de las redes neuronales RBF, que [22] compara con una red ADALINE, sigue el entrenamiento supervisado estándar.

Esto se resume al abordaje realizado en la sección anterior (modelo de regresión logística ordinal basado en los factores) haciendo previamente un análisis factorial y usando después los factores en la regresión logística, con la gran diferencia que la red logra captar complejas relaciones no lineales entre las variables [23]. La superioridad de las RNAs es atribuida también a su mayor parsimonia y, principalmente, a su capacidad de separar los efectos combinados entre las variables de entrada por el paralelismo de su estructura [23], evitando así los problemas de multicolinealidad enfrentados en los modelos clásicos. La RNA 4, de arquitectura RBF 24-8-4, fue la que presentó el mejor aprendizaje (Cuadro



**Figura 2.** Desempeños de los diferentes modelos en la clasificación de madera aserrada de *Pinus caribaea*.

**Figure 2.** Performance of the different models in the classification of *P. caribaea* sawnwood.



**Figura 3.** Importancia de las variables predictoras de la calidad de madera aserrada de *P. caribaea*.

**Figure 3.** Importance of the predictors of *P. caribaea* lumber quality.

**Cuadro 12.** Estadísticas de entrenamiento de las RNAs en la clasificación de la madera aserrada.**Table 12.** ANNs training statistics in sawnwood classification.

RNAs	Arquitectura	Funciones de activación			R (%)	
		Oculto	Salida	Entrenamiento	Prueba	Validación
RNA1	RBF 24-28-4	Gaussian	Softmax	81,420	84,210	68,421
RNA2	RBF 24-32-4	Gaussian	Softmax	76,503	78,947	65,789
RNA3	RBF 24-28-4	Gaussian	Identity	90,710	84,210	84,210
RNA4	RBF 24-8-4	Softmax	Identity	99,453	97,368	98,661
RNA5	RBF 24-29-4	Gaussian	Softmax	53,552	63,158	50,000

**Cuadro 13.** Validación de la red neuronal RBF 24-8-4 en la clasificación de la madera aserrada.**Table 13.** Validation of the RBF 24-8-4 neural network in sawnwood classification.

De \ a	SELECTA	CLASE I	CLASE II	CLASE III	% Aciertos
SELECTA	4	0	0	0	100,00
CLASE I	0	33	5	0	86,84
CLASE II	0	1	5	0	83,33
CLASE III	0	0	1	18	94,74
Porcentaje	5,97	50,75	16,42	26,87	89,55

12) y la mejor capacidad de generalización (Cuadro 13 y Figura 2).

Los resultados de la Figura 2 confirman la significativa mejora que el análisis factorial ha traído al desempeño del modelo de RLMO propuesto y la superioridad del desempeño de las RNAs. Los porcentajes generales de aciertos fueron de 54,44 %, 84,55 % y 89,55 % para los modelos RLMO, RLMO (precedida de factorial) y RBF 24-8-4, respectivamente (Cuadros 4, 11 y 13). Además, la RNA RBF 24-8-4 presentó un porcentaje de sobrevalores (clasificaciones en clases superiores a la correspondiente) de 2,98 %, el cual es inferior al 5%, límite generalmente adoptado para la seguridad de los consumidores.

El análisis de la importancia o contribución de las variables predictoras (Figura 3) indicó que las más importantes fueron: CL (curvatura del Largo), DRch (diámetro de las grietas) y LRch (ancho de las grietas) relacionados con los defectos de secado de la madera aserrada; PERP (podredumbre externa reducida a polvo), DM (daños mecánicos) y DFibr (desviación de las fibras) relacionados con la consistencia de la madera aserrada, MFC (manchas de hongos en el cerne) relacionada con la estética y la gema (G) que es un defecto de aserrado de madera que afecta también la estética.

El resultado anterior indica que medidas deben ser tomadas para controlar, principalmente, los defectos

de secado, consistencia y estética ya que son los más influyentes en la calidad de la madera aserrada. La mayoría de las variables pertenecientes a estos factores caracteriza defectos que aparecen con el tiempo cuando determinados cuidados no se observan en el almacenamiento de la madera aserrada.

## Conclusiones

La regresión logística ordinal precedida de un análisis factorial fue viable para la modelación de la clasificación de la madera aserrada.

La ecuación de regresión logística polinómica ordinal (con los factores) auxilió eficientemente la clasificación de madera aserrada de *Pinus caribaea* Morelet var. *caribaea* Barr. & Golf. en el aserrío Combate de Tenerías de la EFI Macurije.

Las Redes Neuronales Artificiales de tipo RBF, con apenas 8 neuronas en una capa oculta, fueron superiores a los modelos de regresión logística polinómica ordinal en la clasificación de madera aserrada.

Las variables de mayor contribución en la clasificación de la madera aserrada fueron: la curvatura en la anchura (CL), el diámetro de las grietas (DRch), la podredumbre

externa reducida a polvo (PERP), la gema (G) y los daños mecánicos (DM).

## Agradecimientos

El presente trabajo fue realizado con el apoyo de la empresa forestal Macurije (Pinar del Río/Cuba), del Programa Estudiantes-Convenio de Posgrado - PEC-PG de CAPES/Brasil, del Programa de Posgrado en Ciencias Forestales de la Universidad Federal Rural de Pernambuco - PPGCF/UFRPE (Recife/Brasil), del Departamento Forestal de Universidad de Pinar del Río - DF/UPR (Pinar del Río/Cuba).

## Referencias

- [1] M.N.S. Abreu, A.L. Siqueira, W.T. Caiaffa, "Regressão logística ordinal em estudos epidemiológicos", *Rev. Saúde Pública*, vol. 43, no. 1, Feb., pp. 1-11, 2009.
- [2] M.F. Arriaga, G.G. Íñiguez, H.M. Esteban, M.I. Bobadilla, "Madera aserrada de gran escuadría para uso estructural y sus particularidades", *Inform. de la Construcción*, vol. 59, no. 506, Jun., pp. 29-41, 2007.
- [3] C. Bishop, *Neural Networks for Pattern Recognition*, Cambridge, UK: Oxford University Press, 1995.
- [4] M.R. Carreira, "Critérios para classificação de peças estruturais de *Pinus* sp.", Ms. tese, Universidade de São Paulo, São Carlos, SP, 2003.
- [5] M.R. Carreira, A.A. Dias, "Classificação visual de coníferas: análise da aplicação do método norte-americano às espécies de *Pinus* sp. plantadas no Brasil", *Scientia Forestalis*, no. 67, pp.78-87, 2005.
- [6] M.R. Carreira, A.A. Dias, "Critérios para classificação visual de peças estruturais de *Pinus* sp.", *Cadernos de Engenharia de Estruturas*, vol. 8, no. 34, pp. 17-43, 2006.
- [7] H.M. Cartwright, *Artificial Neural Networks*, New York: 2nd ed. Vol 1260 in the Methods in Molecular Biology series, Springer, 2015.
- [8] W.W. Cooley, P.R. Lohnes, *Multivariate data analysis*, NY: John Wiley & Sons, 1971.
- [9] I. Durbak, D.W. Green, T.L. Highley, J.L. Howard, D.B. McKeever, R.B. Miller, R.C. Pettersen, R.M. Rowell, W.T. Simpson, K.E. Skog, R.H. White, J.E. Winandy, J.I. Zerbe, "Wood", In: Kirk-Othmer Encyclopedia of chemical technology, 4th ed., Vol. 25. John Wiley and Sons, Inc., pp. 627-664, 1998.
- [10] C.M. Friel, "Notes on Factor Analysis. Criminal Justice Centre", Texas: Sam Houston State University, 2009. [Online]. Available: [https://www.researchgate.net/profile/Anantjyoti\\_Badgaiyan/post/Any\\_advice\\_on\\_the\\_conflict\\_between\\_results\\_of\\_Exploratory\\_and\\_Confirmatory\\_Factor\\_Analysis/attachment/59d622c579197b8077981164/AS:303652726607873@1449407859829/download/Factor+Analysis.pdf](https://www.researchgate.net/profile/Anantjyoti_Badgaiyan/post/Any_advice_on_the_conflict_between_results_of_Exploratory_and_Confirmatory_Factor_Analysis/attachment/59d622c579197b8077981164/AS:303652726607873@1449407859829/download/Factor+Analysis.pdf)
- [11] D.J.M. García, "Clasificador de madera aserrada. Especies de coníferas. Especificaciones." Ciudad de La Habana: Primera edición, NORMA RAMAL NRAG, Instituto de Investigaciones Forestales, 2006.
- [12] M. Gava, "Viabilidade técnica e econômica da produção de componentes para habitação social utilizando madeira serrada de pinus de terceira classe de qualidade", *Mestrado em Arquitetura e Urbanismo, Universidade de São Carlos, São Carlos, Brasil*, 2005.
- [13] E.A. Godoy, "Aprovechamiento del aserrín mediante su transformación hidrolítica y como fuente de biomasa Cuba", *Revista forestal centroamericana (CATIE)*, vol. 39, no. 40, Dic., pp. 67-69, 1999.
- [14] J.F. Hair, W.C. Black, B.J. Babin, R.E. Anderson, *Multivariate Data Analysis*, NJ: Pearson Prentice Hall, 2010.
- [15] D.E. Kline, P.A. Araman, C. Surak, "Evaluation of an Automated Hardwood Lumber Grading System", In *Proc. The ninth international conference on scanning technology and process optimization for the wood industry*, Seattle, Washington, U.S.A. 2001, pp.141-151.
- [16] P. Klinkhachorn, J. Moody, P. Araman, "Automated lumber processing system: grading the hardwood lumber?" In *Proc. Hardwood Symposium*, 05', 1995, pp. 77-83.
- [17] D.E. Kretschmann, "Commercial Lumber, Round Timbers, and Ties". In: *Wood handbook—Wood as an engineering material: Chapter 6. General Technical Report FPL-GTR-190*, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, 2010, pp.146-170.
- [18] R. Louban, "Theoretical Basis of Adaptive Algorithms with Numerous Practical Applications" in *Image Processing of Edge and Surface Defects*, Springer Series in Materials Science 123, 2009, pp. 174.
- [19] ONEI (Oficina Nacional de Estadística e Información), *Anuario estadístico de Cuba*, La Habana, Cuba, 2014.
- [20] B. Tabachnick, L. Fidell, *Using multivariate analysis*, 5th. Ed. Needham Heights: Allyn & Bacon, 2007.
- [21] E. Thomas, "An artificial neural network for real-time hardwood lumber grading", *Computers and Electronics in Agriculture*, vol. 132, Jan., pp. 71-75, 2017.
- [22] M. Valença, *Fundamentos das redes neurais: Exemplos em Java*, 2ª. Edição, Revisado e ampliado, Olinda-PE: Livro Rápido, Olinda, 2010.
- [23] K. Viriyametantont, "Reconnaissance physique et géométrique d'éléments en béton armé par radar et réseaux de neurones artificiels", *Thèse de Doctorat, Université de Toulouse, France*, 2008.
- [24] X. Xiao, "A multiple sensors approach to wood defect detection", *Ph.D. Dissertation, Virginia Polytechnic Institute, Blacksburg, Virginia*. 1998.