



## ANÁLISIS PRELIMINAR DE LA POTENCIALIDAD DEL COLOR DE LA MADERA COMO VARIABLE PREDICTIVA PARA IDENTIFICACIÓN DE MATERIALES DEL GÉNERO *Eucalyptus*

José BAROTTO<sup>1,2</sup>, María Laura TONELLO<sup>1,2</sup>, Juan Marcelo GAUNA<sup>2,3</sup>, Natalia RAFFAELI<sup>1,2</sup>

### RESUMEN

El objetivo de este trabajo fue analizar la potencialidad predictiva del color de la madera para la identificación de materiales del género *Eucalyptus* implantados en Argentina, mediante el desarrollo de modelos basados en algoritmos de aprendizaje automático. Se trabajó con 15 muestras de 20 materiales de *Eucalyptus* de entre 6 y 7 años, con diferente grado de mejora genética (especies y clones monoespecíficos e híbridos), a los cuales se les determinó el color en el espacio CIEL\*a\*b\*, como también los valores de absorbancia en el rango del espectro visible (400-700 nm). Se probaron dos algoritmos de aprendizaje automático: random forest (RF) y support vector machines (SVM), y se definió como objetivo de la predicción a las especies o híbridos. El modelo desarrollado a partir de RF y los parámetros de color CIEL\*a\*b\* presentó el mejor ajuste, con una exactitud del modelo del 59,7% y un valor de kappa de 0,514, con lo cual puede ser catalogado como moderado. Resulta necesario profundizar este análisis en vista de generar protocolos que permitan su implementación a otra escala, analizando otras secciones de la madera y bajo diferentes tratamientos superficiales.

**Palabras clave:** *eucaliptos*, *CIELab*, *aprendizaje automático*, *Random Forest*, *Support Vector Machines*

### 1. INTRODUCCIÓN

En general, la identificación de especies es un proceso laborioso que requiere de mucho tiempo, y para lo cual suele ser necesario cuantificar características anatómicas y estructurales de la madera (Vieira et al., 2022). Estas características deben ser contrastadas con bases de datos que incluyan esta información, las cuales muchas veces no existen, sumado al hecho de que estas características presentan cierto grado de plasticidad fenotípica (Bambil et al., 2020). Es por lo que se torna necesario el desarrollo de técnicas de identificación alternativas, entre las que puede mencionarse el color de la madera. Sin embargo, la clasificación visual por sí sola es altamente subjetiva y puede ser influenciada por diversos factores, como por ejemplo el nivel de iluminación. El uso de colorimetría puede ser una alternativa válida, debido a su rapidez y facilidad de uso (Vieira et al., 2022). En este sentido, Vieira et al. (2019) mostraron la potencialidad de esta técnica para distinguir especies de mirtáceas nativas de Brasil.

Actualmente, existe una gran variedad de algoritmos de clasificación de aprendizaje automático (machine learning), los cuales pueden aplicarse a un mismo set de datos y comparar los resultados en cuanto a su rendimiento y eficacia del proceso clasificatorio (Bambil et al., 2020). Dos de los más ampliamente utilizados son random forest (RF) y support vector machines (SVM). RF clasifica en función de un conjunto de árboles de decisión, mediante la integración de múltiples modelos para resolver el mismo problema, lo cual aumenta la precisión al disminuir la varianza, y la clasificación se define mediante votación, siendo la más común la votación por mayoría (se asigna la categoría asignada por la mayoría de los clasificadores). SVM pertenece a una familia de clasificadores lineales

<sup>1</sup> Laboratorio de Investigaciones en Madera (LIMAD), Facultad de Ciencias Agrarias y Forestales, Universidad Nacional de La Plata. Contacto: jose.barotto@agro.unlp.edu.ar

<sup>2</sup> Cátedra de Industrias de Transformación Química, Facultad de Ciencias Agrarias y Forestales, Universidad Nacional de La Plata.

<sup>3</sup> Instituto de Fisiología vegetal (INFIVE), Facultad de Ciencias Agrarias y Forestales, Universidad Nacional de La Plata.



no paramétricos, que busca maximizar la distancia que separa los elementos de dos clases diferentes, mediante la definición de un hiperplano óptimo para separar el conjunto de datos en un número discreto de clases predefinidas (Sheykhmousa et al., 2020). Estos y otros algoritmos de aprendizaje automático se han aplicado satisfactoriamente para la identificación de especies a partir de imágenes de hojas (Bambil et al., 2020), árboles (Hu et al., 2018) y de la madera (p. ej.: Rosa da Silva et al., 2022).

El objetivo de este trabajo fue analizar la potencialidad predictiva del color de la madera para la identificación de materiales del género *Eucalyptus* implantados en Argentina, mediante el desarrollo de modelos basados en algoritmos de aprendizaje automático.

## 2. MATERIALES Y MÉTODOS

**Material vegetal, muestreo y acondicionamiento:** se trabajó con material vegetal provisto por la Estación Experimental Agropecuaria (EEA) INTA Concordia y por el Instituto de Recursos Biológicos (IRB) del Centro de Investigaciones Agropecuarias INTA Castelar, y los materiales contaban con 7 y 6 años al momento del apeo, respectivamente (Cuadro 1). Se obtuvieron rodajas a la altura del DAP de diferente número de individuos, representativos del material considerado, que luego fueron procesadas para obtener cubos de aproximadamente 2 cm de lado, tanto de la albura como del duramen, utilizados para las evaluaciones de laboratorio. Se seleccionaron 15 cubos por cada material evaluado, representativos del rango de coloración observado y, a fin de obtener una superficie lo más homogénea posible, fueron lijados en una de sus caras transversales hasta una granulometría de 120. Se seleccionó esta granulometría por cuestiones de disponibilidad de tiempo y la naturaleza preliminar de este trabajo.

**Cuadro 1.** Origen y edad de los materiales evaluados

Origen	Materiales	Edad
INTA Concordia	<u>Especies:</u> <i>E. grandis</i> HSP (EG), <i>E. dunnii</i> (ED), <i>E. benthamii</i> (BE), <i>E. cloeziana</i> (CL), <i>E. camaldulensis</i> (EC) y <i>E. tereticornis</i> (ET). <u>Clones:</u> <i>E. grandis</i> : EG INTA 36 (EG36), <i>E. grandis</i> x <i>E. camaldulensis</i> : GC INTA 8 (GC8), GC INTA 9 (GC9), GC INTA 12 (GC12), GC INTA 24 (GC24), GC INTA 27 (GC27). <i>E. grandis</i> x <i>E. tereticornis</i> : GT INTA 31 (GT31).	7 años
INTA Castelar	<u>Especies:</u> <i>E. sideroxylon</i> (ES), <i>E. viminalis</i> (EV). <u>Clones:</u> <i>E. globulus</i> : GL506, GL515, GL516, GL520, GL526.	6 años

**Determinación de color:** El color se determinó sobre la cara transversal acondicionada, a la cual se le eliminó todo resto de polvo en la superficie de manera previa a la determinación de color. Se realizó una determinación por cubo ( $n = 15$  por material genético), mediante un espectrofotómetro BYK Gardner Color-Guide 45/0, en el rango espectral entre 400 y 700 nm (visible), con iluminante D65, ángulo de observación de  $10^\circ$  y una apertura de 4 mm de diámetro, y estas lecturas se tomaron a una temperatura ambiente de  $24^\circ\text{C}$  y humedad relativa en torno al 80%. Se recolectaron los datos de color en el sistema CIEL\*a\*b\* 1976, en particular  $L^*$  (eje blanco/negro),  $a^*$  (eje rojo/verde),  $b^*$  (eje amarillo/azul) y  $dE^*$  (diferencia de color de la muestra respecto al patrón de referencia), como también con valores de absorbancia en el rango del espectro visible (400-700 nm), registrados cada 20 nm.

**Desarrollo de los modelos y análisis estadístico:** se probaron dos algoritmos de aprendizaje automático: random forest (RF) y support vector machines (SVM), y se definió como objetivo de la predicción a las especies o híbridos. En este sentido, los materiales de *E. grandis* de semilla y el clon EG36 fueron considerados como un solo material ( $n=30$ ), al igual que los 5 materiales híbridos de *E. grandis* x *E. camaldulensis* ( $n=75$ ) y los 5 clones de *E. globulus* ( $n=75$ ). A fin de lograr una representación adecuada, se generó un set de entrenamiento compuesto por el 75% de los cubos de cada material, seleccionados aleatoriamente. A partir de este set se entrenó el modelo, utilizando validación cruzada, con 4 particiones y 100 repeticiones, el cual fue validado frente al 25% de las muestras restantes. Se registró la sensibilidad de los modelos, que refleja el grado de acierto en la predicción de cada material (Aciertos / (Aciertos + Falsos negativos)), así como la exactitud (Aciertos / Total) y el parámetro kappa de Cohen (otra medida de acierto, más adecuada para datos desbalanceados, como en este caso). Todos los



análisis se llevaron a cabo mediante el software R (R Core Team, 2024), y la librería específica *caret*.

### 3. RESULTADOS Y DISCUSIÓN

En el cuadro 2 se detallan los resultados obtenidos por ambos algoritmos de clasificación para cada sistema de color analizado. Es posible observar que el modelo desarrollado a partir de RF para el sistema de color CIEL\*a\*b\* presentó el mejor ajuste de los 4 modelos analizados (2 algoritmos x 2 sistemas de color), con una exactitud del modelo del 59,7% y un valor de kappa de 0,514. En el otro extremo, el modelo desarrollado por SVM para los datos de absorbancia en el espectro visible presentó el ajuste más deficiente (exactitud de 46,3% y kappa de 0,303).

**Cuadro 2.** Exactitud y parámetro Kappa para los dos algoritmos de clasificación ensayados y para ambos sistemas de color probados.

Algoritmo	Sistema de color	Exactitud	Kappa
Support Vector Machines (SVM)	CIEL*a*b*	49,3%	0,339
	Espectro visible	46,3%	0,303
Random Forest (RF)	CIEL*a*b*	59,7%	0,514
	Espectro visible	49,3%	0,357

En el cuadro 3 se detallan los valores de sensibilidad media de los 4 modelos para cada material analizado, así como el desglose de este parámetro en función del algoritmo (promedio de sensibilidad del algoritmo para ambos sistemas de color) y el sistema de color utilizado (promedio de sensibilidad de ambos algoritmos para cada sistema de color). La sensibilidad media de los 4 modelos nos indica, cuáles fueron los materiales que presentaron un mayor grado de acierto en la clasificación por parte de los modelos. Respecto a este indicador, se destaca, en primer lugar, *E. globulus* (GL), con un 82%, seguido por los híbridos GC con un 57%.

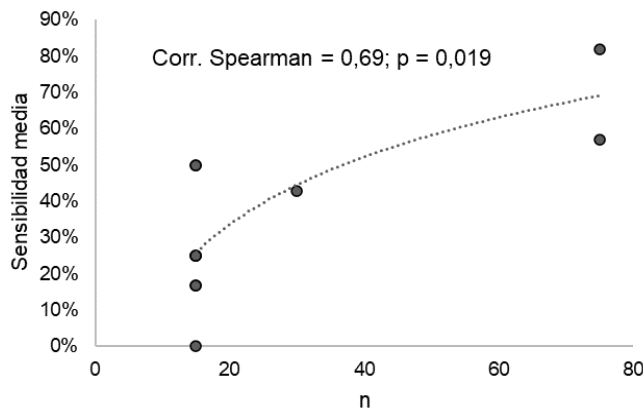
**Cuadro 3.** Sensibilidad media por material, a nivel general y desglosado por algoritmo y sistema de color.

Material	n	Sensibilidad media	Algoritmo		Sistema de color	
			SVM	RF	CIEL*a*b*	Espectro visible
BE	15	25%	0%	50%	33%	17%
CL	15	0%	0%	0%	0%	0%
EC	15	50%	50%	50%	33%	67%
ED	15	17%	0%	33%	17%	17%
EG	30	43%	36%	50%	50%	36%
ES	15	50%	67%	33%	67%	33%
ET	15	25%	17%	33%	17%	33%
EV	15	25%	0%	50%	50%	0%
GC	75	57%	47%	67%	58%	56%
GL	75	82%	94%	69%	83%	81%
GT	15	17%	0%	33%	33%	0%

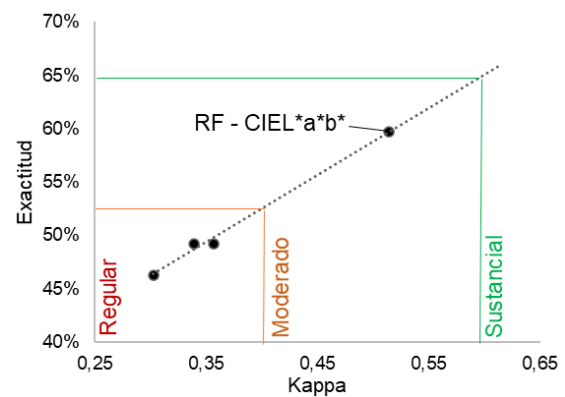
En el otro extremo, *E. cloeziana* (CL) fue el material con menor nivel de predicción, ya que ninguna de las probetas testeadas pudo ser correctamente clasificada. Respecto a este punto, en el Gráfico 1 es posible observar que existe una relación directa y significativa entre el número de probetas analizadas y el grado de sensibilidad en la clasificación. Esto nos indicaría que, para obtener modelos más robustos, es necesario incrementar el número de muestras, lo cual es un resultado lógico y esperable. Si observamos el análisis de sensibilidad media por algoritmo, vemos que RF presentó un nivel de ajuste general superior a SVM, así como también para la mayoría de los materiales analizados. A nivel del sistema de color utilizado, el ajuste fue mejor a partir de los parámetros del sistema CIEL\*a\*b\*. Esto es coincidente con lo reportado en el Gráfico 2, donde se observa una relación directa entre los valores



del parámetro kappa y la exactitud del modelo para los 4 modelos analizados, destacándose como mejor modelo el que fue desarrollado a partir de RF y los parámetros del sistema de color CIEL\*a\*b\*. De acuerdo con la clasificación establecida por Landis y Koch (1977) este modelo puede ser catalogado como moderado, mientras que los restantes entran en la categoría de modelos regulares.



**Gráfico 1.** Sensibilidad media en función del número de muestras de cada material.



**Gráfico 2.** Relación entre el parámetro kappa y la exactitud, y clasificación de los modelos.

#### 4. CONCLUSIONES

El presente trabajo permite establecer el potencial de los parámetros de color de la madera como variables predictivas para la identificación de materiales del género *Eucalyptus*. En este sentido, el modelo desarrollado a partir del algoritmo random forest y los parámetros de color del sistema CIEL\*a\*b\* presentó un grado de ajuste moderado y alta capacidad predictiva, al menos para ciertos materiales. Se destaca la necesidad de incrementar el número de muestras para lograr modelos más robustos. Finalmente, y en vistas de generar protocolos que permitan su implementación a otra escala, se requiere profundizar este trabajo, analizando otras secciones de la madera y bajo diferentes granulometrías de lijado y/o tratamientos superficiales.

#### 5. LITERATURA CITADA

- BAMBIL, D., PISTORI, H., BAO, F., WEBER, V., ALVES, F. M., GONÇALVES, E. G., Figueiredo, L. F. A.; Abreu, U. G. P.; Arruda, R.; BORTOLOTTI, I. M. 2020. Plant species identification using color learning resources, shape, texture, through machine learning and artificial neural networks. *Environment Systems and Decisions*. DOI:10.1007/s10669-020-09769-w
- ROSA DA SILVA, N., DEKLERCK, V., BAETENS, J.M., VAN DEN BULCKE, J., DE RIDDER, M., ROUSSEAU, M., BRUNO, O.M., BEECKMAN, H., VAN ACKER, J., DE BAETS, B., VERWAEREN, J. 2022. Improved wood species identification based on multi-view imagery of the three anatomical planes. *Plant Methods* 18(1): 79. DOI: 10.1186/s13007-022-00910-1.
- HU, M., FEN, H., YANG, Y., XIA, K., & REN, L. 2018. Tree Species Identification Based on the Fusion of Multiple Deep Learning Models Transfer Learning. 2018 Chinese Automation Congress (CAC), Xi'an, China, pp. 2135-2140. DOI:10.1109/cac.2018.8623484
- LANDIS, J.R., KOCH, G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1): 159-174. DOI:10.2307/2529310
- SHEYKHOUSA, M., MAHDIANPARI, M., GHANBARI, H., MOHAMMADIMANESH, F., GHAMISI, P., HOMAYOUNI, S. 2020. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6308-6325. DOI: 10.1109/JSTARS.2020.3026724.
- VIEIRA, H.C., LOPES DA SILVA, E., DOS SANTOS, J.X., BOLZON DE MUÑOZ, G.I., RIBEIRO MORRONE, S., NISGOSKI, S. 2019. Wood colorimetry of native species of Myrtaceae from a Araucaria Forest. *Floresta* 49(2): 353-362. DOI: 10.5380/uf.v49 i2.58236



VIEIRA, H.C., DOS SANTOS, J.X., SOUZA, D.V., RIOS, P.DEA., BOLZON DE MUÑIZ, G.I., RIBEIRO MORRONE, S., NISGOSKI, S. 2022. Applying colorimetry for wood differentiation of fabaceae species grown in southern Brazil. *Maderas. Ciencia y tecnología*, 24(16). DOI: 10.4067/s0718-221x2022000100416