

Detección de especies maderables mediante sensores químicos de olor, aplicando regularización L^1 y modelos de mezclas gaussianas

Wood species detection from chemical array sensors by using L^1 regularization and Gaussian Mixture Models

Detecção de espécies de madeira a partir de sensores de matriz química usando regularização L^1 e modelos de mistura gaussiana

Naren Arley Mantilla Ramírez^a | Iván Darío Porras Gómez^b | Alexander Sepúlveda-Sepúlveda^{*c}

a <https://orcid.org/0000-0002-3185-7387> Universidad Industrial de Santander, Bucaramanga, Colombia

b <https://orcid.org/0000-0001-7666-1786> Universidad Industrial de Santander, Bucaramanga, Colombia

c <https://orcid.org/0000-0002-9643-5193> Universidad Industrial de Santander, Bucaramanga, Colombia

- Fecha de recepción: 2022-07-01
- Fecha concepto de evaluación: 2022-09-09
- Fecha de aprobación: 2022-10-19

<https://doi.org/10.22335/rlct.v15i1.1642>

Para citar este artículo/To reference this article/Para citar este artigo: Mantilla Ramírez, N. A., Porras Gómez, I. D., y Sepúlveda-Sepúlveda, F. A. (2023). Detección de especies maderables mediante sensores químicos de olor, aplicando regularización L^1 y modelos de mezclas gaussianas. *Revista Logos Ciencia & Tecnología*, 15(1), 8-18. <https://doi.org/10.22335/rlct.v15i1.1642>

RESUMEN

La identificación de las especies maderables ayuda a combatir el tráfico de especies prohibidas. Una forma de realizarla es mediante su olor. En el presente trabajo se muestra un método que tiene por objetivo la detección de especies maderables a partir de trozos previamente cortados y almacenados, mediante la interacción de los compuestos volátiles que estos emanan con una nariz electrónica. El presente dispositivo procesa la respuesta del arreglo de sensores químicos a través de la regularización lineal y modelos probabilísticos con un enfoque similar al usado en sistemas biométricos. En particular, el método presentado incluye una etapa de estimación de parámetros heurísticos sobre las señales entregadas por el arreglo de sensores, seguido de una etapa de selección de variables por medio de regularización L^1 , para finalmente utilizar modelos de mezclas Gaussianas [GMMs] en el proceso de modelado probabilístico. A modo de resultado, se obtiene un EER promedio de 17.5 % en desempeño en la detección de cuatro especies maderables; y, para el caso particular del Sapán, un EER de 12 % como resultado. En conclusión, este enfoque biométrico da buenos resultados respecto a trabajos anteriores, teniendo en cuenta que en el presente los experimentos se realizan en condiciones más cercanas a la realidad.

Palabras clave: aprendizaje estadístico, reducción de dimensionalidad, modelos de mezclas, narices electrónicas, industria de la madera.

ABSTRACT

Species identification from timber helps combat illegal wood trafficking. In the present work, it is shown a method that aims to detect wood species from previously cut and stored pieces through

the interaction of the volatile compounds that they emanate with an array of chemical sensors (electronic nose). The device processes the response of the chemical sensor array using linear regularization and probabilistic machine learning methods trying to resemble current biometric systems. In particular, this method includes a heuristic parameter estimation stage on the signals provided by the sensor array, followed by a variable selection stage through L^1 regularization, to finally use Gaussian mixture models [GMMs] for the process of probabilistic modeling. As a result, a performance (measured by EER) of 17.5 % is obtained for the detection of four wood species; and, for the particular case of Sapán, an EER of 12 % is achieved. In conclusion, this biometric approach gives good results compared to previous works, taking into account that in the present work, the experiments are carried out in conditions that are closer to reality.

Keywords: statistical learning, dimensionality reduction, mixture models, electronic noses, wood industry.

RESUMO

A identificação das espécies da madeira ajuda a combater o tráfico ilegal de madeira. No presente trabalho, é apresentado um método que visa detectar espécies de madeira de peças previamente cortadas e armazenadas através da interação dos compostos voláteis que elas emanam com um conjunto de sensores químicos (nariz eletrônico). O dispositivo processa a resposta da matriz de sensores químicos usando regularização linear e métodos probabilísticos de aprendizado de máquina tentando se assemelhar aos sistemas biométricos atuais. Em particular, este método inclui uma etapa de estimação de parâmetros heurísticos nos sinais fornecidos pela matriz de sensores, seguida de uma etapa de seleção de variáveis através da regularização L1, para finalmente usar modelos de mistura gaussiana [GMMs] para o processo de modelagem probabilística. Como resultado, obtém-se um desempenho (medido pelo EER) de 17,5% para a detecção de quatro espécies de madeira; e, para o caso particular de Sapán, atinge-se uma EER de 12 %. Em conclusão, esta abordagem biométrica apresenta bons resultados face a trabalhos anteriores, tendo em conta que no presente trabalho as experiências são realizadas em condições mais próximas da realidade.

Palavras-chave: aprendizado estatístico, redução de dimensionalidade, modelos de mistura, narizes eletrônicos, indústria madeireira.

Introducción

La deforestación pone en peligro especies vulnerables como la caoba, el comino crespo, el abarco, el cedro y el sapán (CAS, 2016; WWF, 2012). El transporte y almacenamiento de este tipo de maderas, típicamente, se hace en conjunto con maderas ampliamente usadas comercialmente y más comunes como el pino, mónico, urapán, eucalipto y caracolí, razón por la cual las autoridades tienen dificultades adicionales al tratar de detectar tipos particulares de madera de interés. En consonancia, un procedimiento común llevado a cabo por la Policía Nacional de Colombia consiste en detener un camión que esté transportando madera para pedirle al conductor un salvoconducto; luego, se verifica la carga de madera. Sin embargo, no se tienen instrumentos de monitoreo apropiados para detectar madera de especies de árboles vulnerables y prohibidas por las disposiciones ambientales colombianas, lo cual dificulta el actuar de las autoridades frente a la explotación ilegal de madera.

Las características macroscópicas (color, textura y olor) se pueden utilizar para verificar de manera rápida la respectiva especie de madera (Wheeler, 1998); sin embargo, esta tarea en Colombia se hace de manera empírica por parte de personal capacitado. También existen métodos precisos basados en análisis taxonómicos y genéticos, en los que se comparan muestras de especies de madera a nivel de secuencias genéticas (Yu *et al.*, 2015). Aunque su confiabilidad es de casi 100 %, estas pruebas son costosas, demoradas y deben ser realizadas por expertos localizados en centros urbanos alejados del lugar donde se realiza el análisis. Otras técnicas utilizadas involucran análisis de espectroscopía (Cabral *et al.*, 2012; Zhao y Jun, 2016), pero que siguen requiriendo el apoyo de expertos y toman bastante tiempo. Además, está el método de análisis anatómico mediante caracterización microscópica (Urrego *et al.*, 2012); no obstante, este enfoque aún requiere de mano de obra especializada, que es costosa y de difícil consecución en lugares remotos. Estos métodos son técnicas efectivas,

pero aún no cumplen los requisitos para ser aplicados en regiones no urbanas y rurales alejadas de las principales ciudades (Kalaw *et al.*, 2018). Otra forma de resolverlo es utilizando técnicas basadas en tratamiento de imágenes (FRIM&UTAR, 2018; Agritix, 2016); a pesar de ello, no todas las especies se pueden identificar de esta manera, especialmente si son de la misma familia (Cordeiro *et al.*, 2016).

De otra parte, soluciones alternativas proponen analizar los compuestos volátiles emitidos por las especies de madera mediante el uso de estrategias como la cromatografía de gases, que permite identificar compuestos volátiles específicos dentro de una mezcla (Fedele *et al.*, 2007); pero, también es una técnica costosa. Una opción bastante más económica y práctica consiste en utilizar sensores químicos de olores (Kalaw *et al.*, 2018). Estos dispositivos (también llamados narices electrónicas) brindarían información complementaria a otras propiedades de las especies de madera. Además, esta estrategia facilitaría a la Policía Nacional detectar aquellas porciones de madera escondidas de entre aquellos trozos para los cuales las autoridades ambientales sí han otorgado el respectivo permiso; y, ello se podría realizar sin requerir descargar parte del pesado cargamento. Aunque existen modelos de narices electrónicas disponibles en el mercado, también es posible fabricarlas formando matrices de sensores químicos de aromas.

Autores como Baietto *et al.* (2010), Wilson (2012; 2013), Baietto *et al.* (2013), Cordeiro *et al.* (2016) y Kalaw *et al.* (2018), reportan el uso de sensores químicos para la identificación de especies maderables a partir de los compuestos volátiles que estas emanan; sin embargo, estos estudios incluyen pocas especies maderables, e. g., Cordeiro *et al.* (2016), pocas observaciones y poca variabilidad en sus datos, e. g., Wilson *et al.* (2005). Para el caso de especies maderables encontradas en Colombia, los trabajos son escasos. En uno de ellos, Mantilla *et al.* (2021) realizan un análisis de componentes principales para luego aplicar una estrategia de clasificación basada en máquinas de vectores de soporte con el fin de identificar especies maderables a través del uso de un prototipo de nariz electrónica, logrando un desempeño de 80 para la tarea de clasificación de entre cuatro especies maderables

diferentes. Ese trabajo recién mencionado, al igual que los trabajos encontrados a nivel mundial, se centra en la tarea de identificar especies, cuando en la práctica, la detección de especies resultaría más útil.

Además, en los trabajos referenciados se realizaron tareas de identificación de especies de madera, lo que implica un número limitado y cerrado de especies dentro de las cuales clasificar una muestra. Como alternativa, se propone un procedimiento de verificación, más útil en la práctica, en el que se compara una muestra con un modelo de referencia correspondiente a una especie de interés. Si la muestra de prueba no se parece a la clase objetivo, con relación a un modelo que representa el resto de especies, se dice que pertenece a alguna otra clase. Como la identificación de maderas es un problema cerrado (hay que escoger entre alguna de las N clases posibles conocidas), este enfoque parece insuficiente para la detección de especies de madera. La detección de especies de madera podría abordarse como un problema abierto, es decir, determinando si la muestra corresponde a una especie de interés particular (asumiendo que se tiene un modelo de referencia de ésta), ó, si tiene alguna otra de identidad desconocida. En lo que se refiere a detección (verificación) de especies maderables, se han utilizado modelo de mezclas gaussianas, obteniendo un EER de 33.64 % (Mantilla *et al.*, 2020).

En el presente trabajo se presenta un método de detección de especies maderables, inspirado en los sistemas biométricos, a partir de las señales entregadas por sensores químicos de sustancias volátiles. El método incluye una etapa de estimación de parámetros heurísticos, seguida de una de selección de variables mediante LASSO.

LASSO [*Least Absolute Shrinkage and Selection Operator*] es una técnica de regularización que implica ajustar un modelo con todos los p predictores, pero agregando restricciones con el fin de obtener modelos menos complejos, esto hace que algunos coeficientes asociados a sus respectivas entradas tiendan a cero (James *et al.*, 2013). Este enfoque puede ser utilizado para la selección de entradas y sensores.

LASSO identifica las variables independientes que afectan significativamente la variable de respuesta agregando un término de regularización con norma \mathbb{L}^1 de la forma $\|\beta\| = \sum_i |\beta_i|$ a la regresión

$$\hat{\beta} = \min_{\beta} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Donde $\beta_0, \beta_1, \dots, \beta_p$ son los coeficientes del modelo dado por

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$$

x_{ij} corresponde al valor de la j -ésima variable de la i -ésima observación, y , $\lambda \geq 0$ es un parámetro de ajuste. Para un valor de λ suficientemente pequeño, el término de regularización tiende a cero y el estimador se convierte en "Mínimos Cuadrados ordinarios"; no obstante, para un valor de λ suficientemente grande, los estimadores β_j son los que tienden a cero (Friedman *et al.*, 2009).

■ Método

Arreglo de Sensores Químicos

El sistema de medición de olores mediante arreglos de sensores químicos utilizado en el presente trabajo corresponde al desarrollado

Tabla 1

Listado de sensores del arreglo de sensores químicos que hacen parte de la nariz electrónica

ID Sensor	Marca	Referencia
1	HANWEI	MQ-2
2	HANWEI	MQ-3
3	HANWEI	MQ-4
4	HANWEI	MQ-6
5	HANWEI	MQ-7
6	HANWEI	MQ-8
7	HANWEI	MQ-135
8	HANWEI	MQ-9
9	FIGARO	TGS-832
10	HANWEI	MQ-6
11	FIGARO	TGS-823
12	FIGARO	TGS-816
13	FIGARO	TGS-822
14	FIGARO	TGS-813
15	FIGARO	TGS-826
16	HANWEI	MQ-3

en Ruiz-Jiménez (2018). Este tipo de sistemas, en general, consta de tres fases: primero, una fase química que transforma la interacción de los compuestos volátiles con el arreglo de sensores de gas en señales eléctricas; segundo, una fase electrónica, que es donde se adquieren las señales eléctricas y estas se acondicionan para obtener una representación matricial de la evolución temporal de la interacción de los sensores con la muestra; y tercero, una fase donde se realiza el procesamiento de los datos obtenidos. Existen diferentes tipos de sensores comerciales, que varían en tamaño, sensibilidad, aplicación y tecnología utilizada, pero los más comunes son aquellos basados en películas semiconductoras de óxido metal. A pesar de que estos sensores requieren de módulo de precalentamiento, ellos son preferidos porque presentan características estables a lo largo del tiempo y son de fácil mantenimiento (Ghasemi-Varnamkhasi *et al.*, 2019).

El arreglo total corresponde a una matriz de $4 \times 4 = 16$ sensores semiconductores de óxido-metal, cuya resistencia eléctrica varía a causa de la reacción química que ocurre cuando los compuestos volátiles hacen contacto con los sensores. En particular, se usan sensores de las casas fabricantes *Figaro Engineering* y *Hanwei Electronics*, que se caracterizan por su bajo costo y por tener capacidad para detectar bajas concentraciones de gas. En la Tabla 1 se listan los sensores que hacen parte de la matriz de sensores.

Representación de los datos de curvas de respuesta de sensores

Se tomaron 309 muestras de variados tipos de maderas en diferentes depósitos de madera localizados en poblaciones de la región del Gran Santander (Colombia): Bucaramanga, Lebrija, Socorro, San Gil, Pamplona y Cúcuta. Los datos fueron tomados a un periodo de muestreo de 270 ms, predefinido en el prototipo usado. Los datos resultantes de este proceso reposan en <https://data.mendeley.com/datasets/gvwr8hmv2t/1>

Diferentes características se pueden extraer de las curvas de respuesta de conductancia para cada sensor. En particular, se reporta el uso de los valores máximo, mínimo y área bajo la curva de respuesta de cada sensor (Yan *et al.*, 2015); y, en otros trabajos se realiza un ajuste a modelos predefinidos (Carmel *et al.*, 2003). En el presente trabajo se utilizaron las siguientes 6 características:

- G_{0r} , valor de conductancia inicial, media de las primeras 100 muestras de la respuesta total.
- G_{fr} , valor de conductancia final, media de las últimas 50 muestras de la curva de respuesta.
- G_{maxr} , valor de conductancia máxima.
- G_{minr} , valor de conductancia mínima.
- B , coeficiente de ganancia y A , localización del polo proveniente del ajuste a un modelo autorregresivo de primer orden:

$$H(z) = \frac{Bz^{-1}}{1 + Az^{-1}}$$

Las primeras cuatro características mencionadas corresponden a parámetros extraídos directamente de cada curva de respuesta, que es una de las formas más comunes en trabajos consultados. Las otras dos características, que corresponden al ajuste a un modelo regresivo de primer orden, son una manera de intentar concentrar la forma total de la curva en solo dos parámetros. En resumen, se extraen seis valores por cada una de las 16 curvas de res-

puesta para cada muestra de madera, con lo que se obtiene un total de 96 características por cada huella odorífica, conformando una matriz $X_{309 \times 96}$ (309 observaciones de dimensión 96).

Este set de datos sigue siendo aún de tamaño reducido. Con la finalidad de sobrellevar este problema, en el presente trabajo se utiliza la técnica SMOTE [*Synthetic Minority Oversampling TEchnique*] (Chawla *et al.*, 2002), que es una técnica de sobremuestreo para nuevas instancias de entre (en medio de) las ya existentes. Para generar un dato sintético, se parte de un dato real y sus k vecinos más cercanos. Se traza una línea (de forma imaginaria) desde el dato real hasta cada uno de sus vecinos y, sobre estas líneas, se escoge aleatoriamente un punto que será el dato sintético. Este procedimiento se realiza para cada dato real y se repite cuantas veces sea necesario hasta obtener el número de instancias sintéticas deseado. El aumento se hizo hasta completar 216 muestras para las cuatro clases con mayor cantidad de muestras, y otras 216 para el resto de muestras del "resto del universo"; con esto se llegó a una matriz de características de tamaño $X_{1080 \times 96}$. En la Tabla 2 se muestra la distribución de muestras por clase.

Tabla 2

Distribución de muestras por especie maderable

	Nombre científico	Cantidad original	Cantidad resultante
Cedro	<i>Cedrela odorata</i>	84	216
Mónico	<i>Cordia gerascanthus</i>	47	216
Pino	<i>Retrophyllum rospigliosii</i>	27	216
Sapán	<i>Clathrotropis brunnea</i>	43	216
Gayacán, tachuelo, urapán, caracolí, cedrillo, abarco, higuierón, roble, trompillo, cafeto, escobo, eucalipto, caoba	<i>Tabebuia aurea, Zanthoxylum rhoifolium, Fraxinus uhdei, Anacardium excelsum, Simarouba amara, Cariniana pyriformis, Ficus spp., Quercus humboldtii, Guarea guidonia, Coffea arabica, Alchornea triplinervia, Corymbia citriodora, Swietenia macrophylla.</i>	108	216

Enfoque biométrico para la verificación de especies de madera

Como la idea principal es apoyar a las autoridades en su lucha contra la tala ilegal y selectiva de especies de madera, los algoritmos de identificación (clasificación) pueden fallar porque eligen dentro de un grupo cerrado de posibilidades; en tanto que, la estrategia de verificación modela en una única función de densidad de probabilidad (*Universal Background Model*, UBM) aquellas especies que no hacen parte del objetivo de búsqueda por parte de las autoridades. La verificación de especies podría ser un mejor enfoque desde el punto de vista práctico.

Este modelo se compara con un modelo de características específicas de una especie de madera, para tomar una decisión de aceptación (hipótesis nula) o rechazo (hipótesis alternativa). La tarea de verificación se puede resumir al probar si una muestra corresponde a la clase analizada (hipótesis nula) o a una clase desconocida (hipótesis alternativa). En este caso, la hipótesis del impostor (cualquier otra clase) está modelada por el *Universal Background Model* (UBM) (Reynolds *et al.*, 2000).

El elemento UBM es básicamente una función de densidad de probabilidad (PDF, *Probability Density Function*) que representa las propiedades de la huella odorífica de la población de especies de referencia. En ese sentido, la huella odorífica *dudosa* (la que se quiere analizar) se comparó con respecto al UBM, así como a un

modelo PDF de una especie de madera en particular. En ese caso, hay dos modelos: el modelo de una especie de madera (λ_s) y el modelo de referencia UBM (λ_0). Al analizar las observaciones correspondientes a la señal interceptada \mathbf{X} , se obtienen dos valores de probabilidad, $P(\mathbf{X}/\lambda_s)$ y $P(\mathbf{X}/\lambda_0)$, con los que se construye la Razón de Verosimilitud (LR, por sus siglas en inglés, *Likelihood Ratio*). Sin embargo, es común usar el Log-Likelihood Ratio (LLR),

$$\ell(\mathbf{X}) = \log P(\mathbf{X}/\lambda_s) - \log P(\mathbf{X}/\lambda_0)$$

A medida que aumenta el valor $\ell(\mathbf{X})$, la evidencia de que la huella odorífica dudosa corresponde a la especie que estamos buscando se vuelve más fuerte.

Para el modelado de la PDF, se prefiere el conocido modelo de mezclas gaussianas (GMM, *Gaussian Mixture Models*). El uso de una mezcla de modelos gaussianos está motivado por su capacidad para modelar densidades arbitrarias (Porrás *et al.*, 2020). Una GMM se compone de una mezcla finita de componentes gaussianos multivariados y el conjunto de parámetros indicados por λ . Se caracteriza por una combinación lineal ponderada de densidades gaussianas unimodales de C mediante la función:

$$P(\mathcal{O}/\lambda) = \sum_{i=1}^C \pi_i \mathcal{N}(\sigma; \mu_i, \sigma_i)$$

Donde \mathcal{O} es una observación o vector de características de dimensión D , π_i es el peso de mezclado (probabilidad anterior) del i -ésimo

componente Gaussiano, y $\mathcal{N}(\sigma; \mu_i, \sigma_i)$ es la función de densidad Gaussiana D -variada con vector media μ_i y matriz de covarianza σ_i . El popular algoritmo de *Expectation Maximization* (EM) se utiliza para maximizar la probabilidad con respecto a un dato dado.

■ Resultados

Resultados de la selección de características usando LASSO

En total se tienen 96 características, 6 parámetros heurísticos estimados (como se describe en la sección 2.2) por cada uno de los 16 sensores químicos. La cantidad de datos corresponde a 1080, lo que genera la matriz de datos $\mathbf{X}_{1080 \times 96}$. Las primeras 16 características (1-16) corresponden al valor de G_0 para los 16 sensores; las siguientes corresponden a los 16 valores de G_f (17-32), los 16 valores de G_{max} (33-48), los valores de G_{min} (49-64), los valores de A (65-72) y, finalmente, los valores de B (73-96).

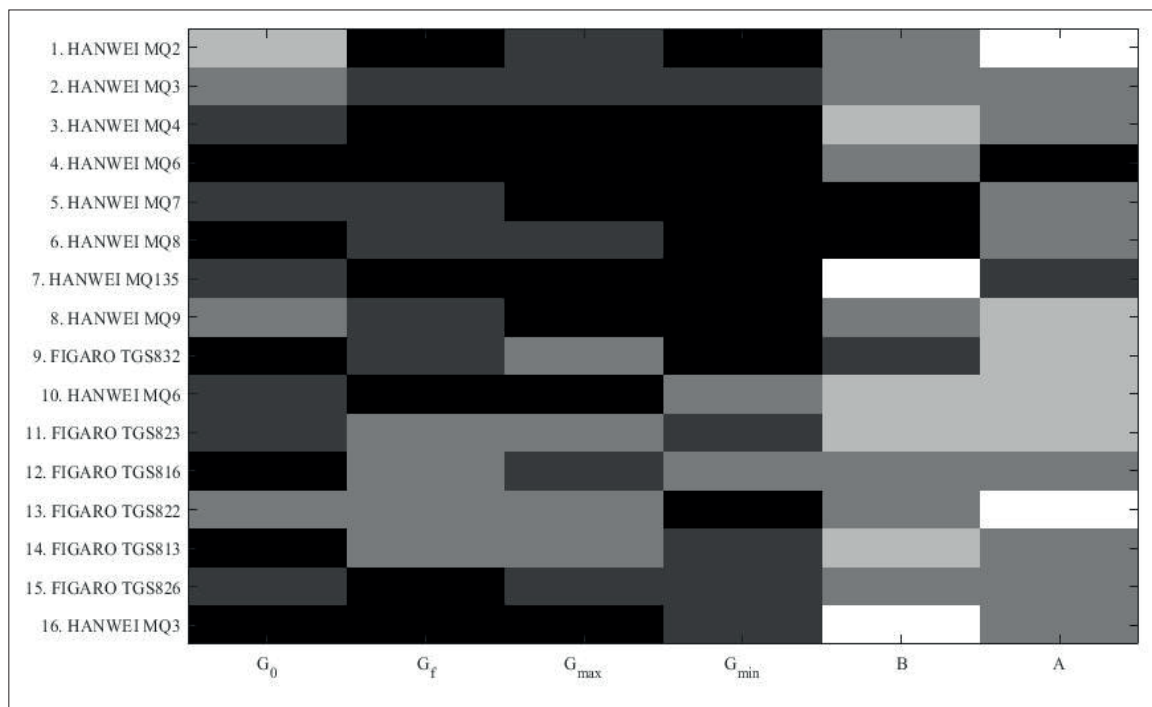
El set aumentado se divide en cinco subgrupos. Cuatro de esos grupos se utilizan para la selección de características con LASSO, y el grupo restante emplea para validar el modelo obtenido en la fase de entrenamiento, basado en las características seleccionadas con LASSO. Esto se hace cinco veces, para que cada subgrupo sea usado como conjunto de validación (validación cruzada, *k-fold*, con $k=5$). Este proceso se realizó 25 veces y, para cada una de las cuatro siguientes especies se repitió el procedimiento: cedro, mónico, pino y sapán. Aquellas características que recurrentemente son seleccionadas por LASSO, son consideradas como las que representan mayor relevancia para la clasificación de esas cuatro especies. En la figura 1 se muestran las características seleccionadas con mayor frecuencia para cada una de las clases.

En la figura 1 se puede observar que las características que representan mayor relevancia son aquellas relacionadas con el modelo autorregresivo, es decir: A y B . Adicionalmente, se puede inferir que las referencias de los sensores que aportan más información para la clasificación de maderas son:

- Sensores 2 y 16: HANWEI MQ-3, que es especialmente sensible a los alcoholes.
- Sensor 8: HANWEI MQ-8, utilizado principalmente para la detección de fugas de gas, por su sensibilidad al gas hidrógeno (H_2).
- Sensor 10: HANWEI MQ-6, que se utiliza principalmente en la detección de fugas de gas, por su alta sensibilidad a gases licuados de petróleo, propano y butano.
- Sensor 11: FIGARO TGS-823, especialmente sensible a vapores de etanol.
- Sensor 12: FIGARO TGS-816, sensor de propósito general con sensibilidad a un amplio rango de gases combustibles como metano, propano y butano.
- Sensor 13: FIGARO TGS-822, altamente sensible a los vapores de solventes orgánicos, así como a otros vapores volátiles. También tiene sensibilidad a gases combustibles como el monóxido (CO) de carbono, por lo que es un buen sensor de uso general.
- Sensor 14: FIGARO TGS-813, altamente sensible al propano, metano y butano, lo que lo hace muy utilizado en aplicaciones relacionadas con detección de fugas de gases.

Figura 1

Distribución de veces que cada característica fue seleccionada por LASSO



Nota. En color claro, aquellas que más se repiten para el caso de clasificación de las especies cedro, mónico, pino y sapán.

De otra parte, a partir de este análisis se pueden identificar aquellos sensores, dentro de los incluidos en el arreglo, que aportan mayor información desde el punto de vista de separabilidad de las clases de maderas en el espacio de las mediciones de los sensores odoríficos. Cabe mencionar que el criterio inicial utilizado para la selección de los sensores que forman el arreglo fue una mezcla entre variedad en el tipo de sensores y disponibilidad de los mismos en el mercado.

Detección de especies de maderas

Para medir el rendimiento del sistema de detección de especies de madera propuesto, se utilizan las curvas DET (*Detection Error Trade-off*, en inglés) y el valor EER (*Equal Error Rate*, en inglés). Las curvas DET trazan la tasa de falso rechazo (FRR, *False Rejection Rate*) en el eje y , versus la tasa de falsa aceptación (FAR, *False Acceptance Rate*) en el eje x , donde la curva más cercana a la esquina inferior izquierda del gráfico corresponde al sistema que tiene la mejor actuación. La tasa de falso rechazo

(FRR) representa la tasa de muestras que serán rechazadas perteneciendo a la clase de interés (también conocidos como falsos negativos); mientras que, la tasa de falsa aceptación representa las muestras que serán aceptadas, aunque, en realidad, pertenecen a la clase conocida (también llamados falsos positivos). El valor EER corresponde a la intersección de una línea de pendiente 1 con la curva DET; es decir, cuando FFR es igual a FAR.

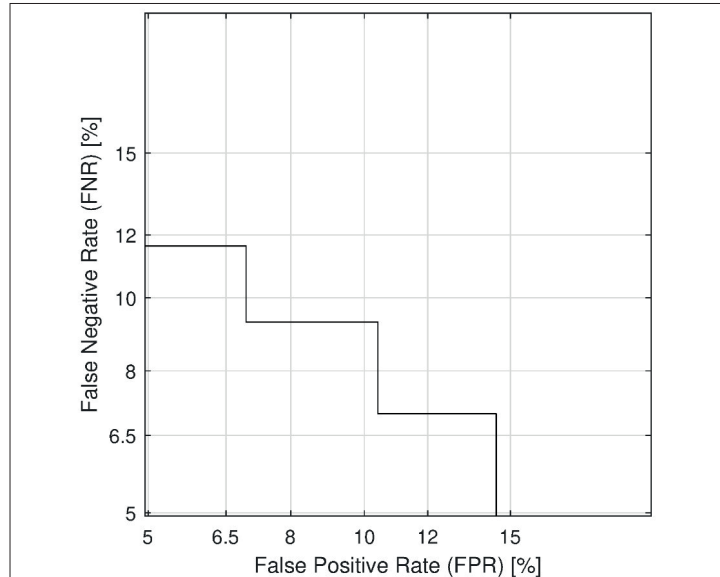
En el proceso de entrenamiento y prueba del modelo se usa un conjunto de datos conformado por 1080 muestras de al menos 18 especies de madera, distribuidos como se muestra en la Tabla 2. 691 muestras de la categoría *otras especies* se usaron para adaptar el UBM correspondiente a un GMM con 4 funciones Gaussianas. Para todos los experimentos, la PDF (*Probability Density Function*, en inglés) de cada clase, que representa una especie de madera particular, está modelada por una mezcla de 4 modelos gaussianos. El valor EER se estima utilizando un procedimiento de validación cruzada de 5 conjuntos; donde el 80 %

de las muestras de madera corresponden al conjunto de entrenamiento, mientras que las muestras restantes (20 %) se utilizan para la validación. Los experimentos de verificación se aplican a partir de las características extraídas con LASSO. Este experimento se repite 100 veces, con diferentes ordenaciones aleatorias del conjunto de datos. La verificación con la

clase cedro, aplicando LASSO, mostró una tasa de error de clasificación de 17.83 %. Al igual que con el cedro, el experimento de verificación se realizó para otras tres especies (móncono, pino y sapán). Un resumen de los resultados se puede ver en la Tabla 3, y en la Figura 2 se muestra una gráfica de la curva DET para el experimento *Detección de Sapán*.

Figura 2

Curva DET (Detection Error Trade-off) para la Detección de sapán, aplicando LASSO como selección de características



Nota. Tasa de falsos negativos (FNR, por sus siglas en inglés). Tasa de falsos positivos (FPR, por sus siglas en inglés).

Tabla 3

Valores de EER (Equal Error Rate) para la detección de especies maderables (Cedro, Móncono, Pino, Sapán) mediante un enfoque biométrico que utiliza LASSO y GMM-UBM

Especie	EER promedio	Error Estándar
Cedro	17.83 %	0.36
Móncono	21.12 %	0.56
Pino	19.05 %	0.17
Sapán	11.92 %	0.33

Al revisar el estado del arte, son muy escasos los trabajos en los que se aborda el problema de detección de maderas mediante olores desde un enfoque biométrico (una excepción sería Mantilla *et al.*, 2020), y, por tanto, los resultados del presente trabajo no son directamente comparables con esos trabajos; además, en trabajos reportados, como en Kalaw *et al.* (2018), Cordeiro *et al.* (2016) y Wilson (2012), típicamente se utiliza material de madera

recién cortada, cuyos aromas son frescos y aún fuertes. En contraste, en el presente trabajo las condiciones experimentales se acercan más a situaciones reales.

Al comparar con trabajos más recientes, el valor EER promedio de 17.5 % (12 % para el caso del Sapán) es notablemente mejor que valores como el de Mantilla *et al.* (2020), donde se reporta un EER de 33.6 %.

■ Conclusiones

En el presente trabajo se responde a la pregunta de si, en efecto, se pueden utilizar sensores químicos de aromas para identificar especies de madera. A modo de aporte, se propuso un enfoque de tipo biométrico para el problema de detección de maderas a partir de olores; es decir, se detectan especies de madera a partir de la huella odorífica de la

especie de madera en cuestión (en analogía a como se verifica la identidad de un hablante a partir de su huella de voz). Este enfoque, para el caso de identificación de especies de madera, es particularmente novedoso ya que cuantifica, con medidas de tipo biométrico, la capacidad de los sensores de olor para detectar especies de madera.

A modo de resultado final, se cuantifica el poder de una matriz de sensores odoríficos para detectar especies de madera. Se observan valores aceptables de desempeño, a pesar de utilizar sensores comerciales de propósito general; sin embargo, a modo de limitación, se tiene que los resultados son solo para unas pocas especies de madera accesibles en aserríos y carpinterías.

En cuanto a trabajos futuros, se propone realizar un análisis de los compuestos volátiles que forman los aromas de las diferentes especies de madera, para así determinar cuáles serían aquellos sensores que mejor se ajusten. No obstante, ello requiere de un análisis químico previo de los aromas presentes en las especies de maderas a detectar. De otra parte, se sugiere considerar estrategias de mezclado de información con el fin de combinar información proveída por la nariz electrónica con información de imágenes acerca del color, o con imágenes de la anatomía microscópica de las especies. Al mezclar ambos tipos de información, complementarias entre sí, se podría obtener un sistema más preciso.

Referencias

- Agritix. (2016). *Xylorix: Macroscopic wood identification system*. <https://www.xylorix.com>
- Baietto, M., Pozzi, L., Wilson, A., y Bassi, D. (2013). Evaluation of a portable MOS electronic nose to detect root rots in shade tree species. *Computers and Electronics Agriculture*, 96, 117-125. <https://doi.org/10.1016/j.compag.2013.05.002>
- Baietto, M., Wilson, A., Bassi, D., y Ferrini, F. (2010). Evaluation of three electronic noses for detecting incipient wood decay. *Sensors*, 10(2), 1062-1092. <https://doi.org/10.3390/s100201062>
- Cabral, E., Simas, R., Santos, V., Queiroga, C., Da Cunha, V., De Sá, G., Daroda, R., y Eberlin, M. (2012). Wood typification by Venturi easy ambient sonic spray ionization mass spectrometry: The case of the endangered Mahogany tree. *Journal of Mass Spectrometry*, 47(1), 1-6. <https://pubmed.ncbi.nlm.nih.gov/22282083/>
- Carmel, L., Levy S., Lancet, D., y Harel, D. (2003). A feature extraction method for chemical sensors in electronic noses. *Sensors and Actuators B: Chemical*, 93(1), 67-76. [https://doi.org/10.1016/S0925-4005\(03\)00247-8](https://doi.org/10.1016/S0925-4005(03)00247-8)
- Chawla, N. V., Bowyer, K. W., Hall, L. O., y Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357. <https://www.jair.org/index.php/jair/article/view/10302>
- Cordeiro, J., Li, R., Takahashi, E., Rehder, G., Ceccantini, G., y Gruber, J. (2016). Wood identification by a portable low-cost polymer-based electronic nose. *RSC Advance*, 6(111). <https://pubs.rsc.org/en/content/articlelanding/2016/ra/c6ra22246c>
- Corporación Autónoma Regional de Santander [CAS] (2016). *En Santander hay 13 especies de flora y 12 de fauna priorizados para su conservación*. <http://cas.gov.co/index.php/sala-de-prensa/453-ensantander-hay-13-especies-de-flora-y-12-de-fauna-priorizados-para-su-conservacion.html>
- Fedele, R., Galbally, I., Porter, N., y Weeks, I. (2007). Biogenic VOC emissions from fresh leaf mulch and wood chips of *Grevillea robusta* (Australian Silky Oak). *Atmospheric Environment*, 41(38), 8736-8746. <https://doi.org/10.1016/j.atmosenv.2007.07.037>
- Forest Research Institute Malasy y University Tunku Abdul Rahman [FRIM&UTAR] (2018). *My wood: Mobile Wood identification for 100 Malaysian timbers*. <http://mywoodid.frim.gov.my>
- Friedman, J., Hastie, T., y Tibshirani, R. (2009). *The Elements of Statistical Learning*. <https://link.springer.com/book/10.1007/978-0-387-84858-7>
- Ghasemi-Varnamkhasti, M., Razdari, A., Yoosefian, S., Izadi, Z., y Rabiei, G. (2019). Selection of an optimized metal oxide semiconductor sensor (MOS) array for freshness characterization of strawberry in polymer packages using response surface method (RSM). *Postharvest Biology and Technology*, 151, 53-60. <https://doi.org/10.1016/j.postharvbio.2019.01.016>
- James, G., Witten, D., Hastie, T., y Tibshirani, R. (2013). *An introduction to statistical learning*. Springer. <https://link.springer.com/book/10.1007/978-1-4614-7138-7>

- Kalaw, J., y Sevilla, F. (2018). Discrimination of wood species based on a carbon nanotube/polymer composite chemiresistor array. *Holzforschung*, 72(3), 215-223. <https://www.degruyter.com/document/doi/10.1515/hf-2017-0097/html>
- Mantilla, N., Ruiz, L., Ortega, H., Paja, M., y Sepúlveda, A. (2020). A low cost electronic nose with a GMM-UBM approach for wood species verification. *ICPRAM-2020*, 333-341. https://scholar.google.com/citations?view_op=view_citation&hl=es&user=Deodh2sAAAAJ&citation_for_view=Deodh2sAAAAJ:d1gkVwhDpl0C
- Mantilla, N., Ruiz, L., Ortega, H., y Sepúlveda, A. (2021). Identificación de especies de maderas locales mediante el uso de nariz electrónica y aprendizaje automático: Un experimento preliminar. *INGE CUC*, 17, 188-200. <https://revistas-cientificas.cuc.edu.co/ingecuc/article/view/2570>
- Porras, D., Sarria, M., y Sepúlveda, A. (2020). Speaker Verification system based on articulatory information from ultrasound recordings. *DYNA*, 87(213). http://www.scielo.org.co/scielo.php?pid=S0012-73532020000200009&script=s_ci_arttext&tlng=en
- Reynolds, D., Quatieri, T., y Dunn, R. (2000). Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10(1-3), 19-41. <https://doi.org/10.1006/dspr.1999.0361>
- Ruiz-Jiménez, L. F. (2018). *Detección de los insectos de la subfamilia Triatominae basado en narices electrónicas*. Technical report, Universidad Industrial de Santander.
- Urrego, D., Pérez, M., Polanco, C. y Pérez, M. (2012). Control del aprovechamiento ilegal de flora a partir de la anatomía de siete especies maderables en estado de amenaza. Aportes a la política nacional ambiental. *Revista Criminalidad*, 54(1), 259-281. http://www.scielo.org.co/scielo.php?script=s_ci_arttext&pid=S1794-31082012000100004
- Wheeler, E. A., y Baas, P. (1998). Wood identification-a review. *IAWA journal- International Association of Wood Anatomists*, 19(3), 241-264. <https://doi.org/10.1163/22941932-90001528>
- Wilson, A., Lester, D., y Oberle, C. (2005). Application of conductive polymer analysis for wood and Woody plant identifications. *For. Ecol. Manage.*, 209(3), 207-224. <https://doi.org/10.1016/j.foreco.2005.01.030>
- Wilson, A. (2012). Application of a conductive polymer electronic-nose device to identify aged woody samples. En S. Yurish, I. Chilibon, V. Carvalho, S. Gervais-Ducouret (eds.). *Proceedings of The 3rd International IARIA Conference on Sensor Device Technologies and Applications, Rome, Italy. Xpert Publishing Services: Wilmington, DE.* (pp. 77-82). <https://www.fs.usda.gov/research/treesearch/45153>
- Wilson, A. (2013). Diverse Applications of Electronic-Nose Technologies in Agriculture and Forestry. *Sensors*, 13, 2295-2348. <https://www.mdpi.com/1424-8220/13/2/2295>
- World Wildlife Fund [WWF] (2017). *15 especies colombianas de árboles amenazados*. <http://www.wwf.org.co/?uNewsID=299073>
- Yan, J., Guo, X., Duan, S., Jia, P., Wang, L., Peng, C., y Zhang, S. (2015). Electronic nose feature extraction methods: A review. *Sensors*, 15(11), 27804-27831. <https://doi.org/10.3390/s151127804>
- Yu, M., Liu, K., Zhou, L., Zhao, L., y Liu, S. (2015). Testing three proposed DNA barcodes for the Wood identification of *Dalbergia odorifera* T. Chen and *Dalbergia tonkinensis* Prain. *Holzforschung*, 70(2), 127-136. <https://doi.org/10.1515/hf-2014-0234>
- Zhao, P., y Jun, C. (2016). Wood species identification using spectral reflectance feature and optimal illumination radian design. *Journal of Forestry Research*, 27, 219-224. <https://doi.org/10.1007/s11676-015-0171-4>