



**Comparación de algoritmos para detección automática  
de vocalizaciones y llamados de especies de diferente grupo taxonómico**

Santiago Taborda Diosa

Proyecto de investigación para optar al título de Ingeniero Electrónico

Asesora

Claudia Victoria Isaza Narváez, PhD.  
Profesora Titular Universidad de Antioquia

Co-asesora

Maria José Guerrero Muriel, M.Sc.

Universidad de Antioquia  
Facultad de Ingeniería  
Ingeniería electrónica  
Medellín  
2023

---

Cita

(Taborda Santiago, 2023)

---

Referencia

Estilo APA 7 (2020)

Taborda Santiago. (2023). *Comparación de algoritmos para detección automática de vocalizaciones y llamados de especies de diferente grupo taxonómico*. Proyecto de investigación. Universidad de Antioquia, Medellín.

---



Sistemas Embebidos e Inteligencia Computacional (SISTEMIC).

Universidad de Antioquia - [www.udea.edu.co](http://www.udea.edu.co)

**Rector:** John Jairo Arboleda Céspedes.

**Decano/Director:** : Julio César Saldarriaga Molina.

**Jefe departamento:** Eduard Emiro Rodríguez Ramírez

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

## **Resumen**

Este trabajo de investigación se enfocó en la selección y evaluación de cinco propuestas de algoritmos y softwares que permiten realizar identificación automática de especies mediante el análisis de grabaciones de audio, específicamente en el contexto de la biodiversidad del bosque húmedo colombiano. Se seleccionaron metodologías tanto supervisadas como no supervisada, y se realizaron diferentes pruebas utilizando grabaciones de campo que abarcan una diversidad de grupos taxonómicos, como aves, mamíferos y anfibios.

Los resultados revelaron grandes diferencias en el desempeño de cada metodología, influenciadas tanto por el grupo taxonómico abordado como por las características individuales de las especies a identificar. Específicamente, Aureas y Raven Pro/Koogu, destacaron como herramientas altamente versátiles, capaces de ajustarse a las diversas especies contempladas en el análisis.

Las conclusiones del estudio enfatizan la importancia de considerar las características específicas de las especies al elegir una metodología y resaltan la necesidad de mejorar la interfaz de usuario para hacer estas herramientas más accesibles a los expertos en biología. Los hallazgos de esta investigación proporcionan información valiosa para los expertos en biología que buscan seleccionar herramientas apropiadas para el monitoreo acústico y el reconocimiento automático de llamadas de especies en estudios de biodiversidad.

**Palabras claves:** Monitoreo acústico pasivo, identificación de múltiples especies, revisión sistemática, modelos de clasificación, metodologías para detección automática.

## Contenido

Resumen.....	3
Contenido.....	4
Lista de figuras.....	5
Lista de Tablas.....	5
Introducción.....	6
Objetivos.....	8
Marco teórico.....	8
Base de datos y sitio de estudio.....	10
Metodología.....	11
Revisión del estado del arte.....	12
Selección de algoritmos.....	13
Pruebas con los algoritmos.....	14
Evaluación de desempeño.....	15
Métricas de evaluación.....	15
Resultados y análisis.....	16
Revisión sistemática del estado del arte.....	16
Selección de software y metodologías.....	17
Descripción de parámetros y pruebas con las metodologías seleccionadas.....	19
- Kaleidoscope Pro:.....	19
- Arbimon Random Forest:.....	22
- Arbimon Clustering (BETA):.....	24
- Aureas:.....	26
- Koogu / Raven Pro:.....	27
- ANIMAL-SPOT:.....	28
Conclusiones.....	31
Referencias Bibliográficas.....	32
Anexos.....	35
Anexo 1: Recopilación de bibliografía.....	35

## Lista de figuras

Figura 1. Mapa de la ubicación del sitio de estudio. Tomado de Restrepo-Rivas, 2022 [28].....	11
Figura 2. Metodología propuesta.....	12
Figura 3. Criterios de selección de algoritmos.....	14
Figura 4. Metodología para pruebas con los algoritmos seleccionados.....	14
Figura 5. Distribución de los artículos por grupo taxonómico que detectan con el método propuesto en el artículo.....	16
Figura 6. Distribución de los artículos por conjunto de grupos taxonómicos trabajados.....	17
Figura 7. Distribución de los artículos que trabajaron con o sin interfaz de usuario.....	17
Figura 8. Selección de parámetros para análisis de clusters en Kaleidoscope Pro.....	20
Figura 9. Ventanas de resultados en Kaleidoscope.....	21
Figura 10. Resultados F1 score kaleidoscope Pro.....	22
Figura 11. Configuración de modelo de random forest en arbimon.....	23
Figura 12. Resultados de F1 score Arbimon con clasificador Random Forest.....	24
Figura 13. Configuración AED arbimon clustering.....	24
Figura 14. Configuración de análisis de clusters en arbimon.....	25
Figura 15. Clusters generados con la configuración mostrada en la figura 14.....	25
Figura 16. Resultados F1 score arbimon clustering.....	26
Figura 17. Resultados F1 Score Aureas.....	27
Figura 18. Resultados F1 Score Koogu.....	28
Figura 19. Compilación de resultados promedio por Taxón para cada Metodología para especies en el espectro audible.....	30
Figura 20. Compilación de resultados promedio por Taxón para las metodologías que trabajan espectro audible y ultrasonido.....	31

## Lista de Tablas

Tabla 1. Ejemplo de recopilación de bibliografía.....	13
---	----

## Introducción

El monitoreo de los ecosistemas y sus especies es esencial para la comprensión y conservación de la biodiversidad [1]. La identificación de especies se ha llevado a cabo mediante la observación directa y el análisis genético. No obstante, esta metodología enfrenta desafíos significativos como la visualización de individuos en lugares con vegetación densa, común en numerosos ecosistemas colombianos. Por lo tanto, es importante contar con alternativas que brinden apoyo en esta tarea, como el monitoreo acústico pasivo, facilitando la adquisición de datos acústicos. En este tipo de monitoreo, el sonido es registrado en diferentes zonas a través de sensores acústicos y sirve como herramienta para evaluar el comportamiento acústico de las especies de animales que habitan allí y se comunican mediante llamados y/o vocalizaciones. A través de este monitoreo es posible entonces, realizar la detección de la presencia de las especies en un lugar y tiempo determinado y con ello contestar diferentes preguntas de índole biológico que permitan identificar el estado de los ecosistemas debido a impactos externos [2].

Con el paso del tiempo, la capacidad de adquisición y mejoras en los sensores acústicos han permitido la recolección de información acústica durante periodos prolongados de tiempo. Esto trae como consecuencia, la generación de grandes volúmenes de registros acústicos que deben ser procesados y analizados por expertos en biología y ecología. El avance en técnicas de aprendizaje de máquina ha sido un gran aliado a la hora de proponer metodologías [3] [4] [5] (algunas incluidas en softwares como Arbimon [6] y SonoChiro [7]), permitiendo el análisis de grandes cantidades de datos de manera automática.

Actualmente, se cuenta con metodologías que permiten la detección automática de llamados/vocalizaciones de especies de animales. En general, estas metodologías siguen una serie de pasos comunes. En primer lugar, se realiza el preprocesamiento de la señal de audio, que consiste en la generación de una representación en tiempo-frecuencia del audio (espectrograma). En algunos casos, especialmente cuando se trabaja con técnicas de aprendizaje de máquina, se realiza la etapa de segmentación de la señal con el fin de aislar los posibles llamados/vocalizaciones de las especies para luego ser analizadas en una etapa posterior. La siguiente etapa consiste en la extracción de características de los segmentos generados. En esta etapa, se busca extraer la información más significativa de las vocalizaciones de las especies animales. Algunas de las características extraídas en esta etapa son la frecuencia, duración, intensidad e intervalo de repetición de los llamados [8]. Una vez extraídas estas características, se utilizan algoritmos de aprendizaje automático para clasificar los llamados de las especies. Estos algoritmos

pueden usar diversas técnicas de clasificación de tipo supervisado, tales como las Redes Neuronales Convolucionales (CNN) [9] [10] [11], Random Forest [12] [13], Support Vector Machine [14] [13] o técnicas de tipo no supervisado como spectral clustering [29], el algoritmo LAMDA – *Learning Algorithm for Multivariate Data Analysis*- [15] [16] y el algoritmo DBSCAN [4].

Cada técnica ofrece diferentes características y ventajas. El uso de una técnica en particular dependerá de la pregunta específica que se busca resolver (e.g. identificación especie-específica, identificación no supervisada, identificación de individuos de una misma especie, identificación de múltiples especies de animales). Debido a esto, la mayoría de las metodologías propuestas para identificación automática de llamados/vocalizaciones de animales son de tipo supervisado, requiriendo etiquetas para realizar una etapa entrenamiento o aprendizaje, limitándose a identificar sólo aquellas especies conocidas a priori. En países con alta biodiversidad como Colombia, donde hay especies aún desconocidas para la ciencia, obtener etiquetas de especies que habitan en determinados ecosistemas puede ser una tarea difícil. Por ello se ha planteado la necesidad de realizar una identificación de tipo no supervisado que no dependa del conocimiento a priori de las especies presentes en los ecosistemas, ni de etiquetas. Este tipo de metodologías permite, a partir de la identificación de patrones en los datos, encontrar sonotipos que pueden estar asociados a especies en un hábitat específico. El uso de los algoritmos diseñados para la identificación automática de especies ha permitido procesar y analizar grandes volúmenes de datos en tiempos relativamente cortos, en comparación del procesamiento y análisis manual por parte de los expertos, quienes ahora se enfocan en analizar los resultados y observar patrones que les permita identificar el estado de amenaza de las especies y proponer planes de acción y monitoreos tendientes a la conservación de la biodiversidad.

Estos algoritmos pertenecientes, en su gran mayoría al área de la inteligencia computacional, permiten detectar y clasificar vocalizaciones de animales de distintos grupos taxonómicos, cada propuesta implica ventajas y desventajas a la hora de realizar el análisis de datos adquiridos por monitoreo acústico pasivo. Para los expertos en biología no es fácil identificar las ventajas técnicas que trae consigo cada algoritmo. En este sentido se hace necesaria una comparación de estos algoritmos que permita determinar cuál alternativa se ajusta mejor al problema que se quiere abordar. Este trabajo de investigación se centrará en realizar un análisis sistemático de la literatura de algoritmos de reconocimiento de llamadas de especies de diferentes grupos taxonómicos. Se seleccionaron las metodologías más relevantes y en estas se analizaron las diferentes etapas que proponen, haciendo énfasis en la etapa de clasificación y los modelos de aprendizaje automático implementados. Se evalúa el desempeño de las

propuestas seleccionadas, se analizan los requerimientos de hardware (e.g., si necesita instalación o es un servicio en la nube) y sus limitaciones en el análisis de frecuencia (espectro audible o ultrasonido).

## **Objetivos**

### **Objetivo general**

Establecer ventajas y desventajas de algoritmos supervisados y no supervisados para el reconocimiento de especies de diferentes grupos taxonómicos en grabaciones de audio, evaluando su desempeño sobre una base de datos de audios recopilados en zonas geográficas que corresponden a bosque húmedo colombiano.

### **Objetivos específicos**

- Seleccionar un conjunto de algoritmos que realicen identificación automática de especies de diferentes grupos taxonómicos para análisis de metodología y pruebas de desempeño en clasificación.
- Implementar algoritmos seleccionados útiles para identificación automática de especies.
- Evaluar el desempeño de cada algoritmo de identificación automática de especies de diferentes grupos taxonómicos analizando audios registrados en campo.

## **Marco teórico**

El monitoreo de las poblaciones biológicas es una herramienta importante para darse una idea del estado de las poblaciones, y así tomar las medidas necesarias para su conservación [17]. El monitoreo acústico pasivo es una de las alternativas propuestas para este propósito, y consiste en recopilar grabaciones de audio que captan la actividad de los paisajes sonoros [18]. Este tipo de monitoreo permite la identificación de especies de una forma menos invasiva sin la presencia constante de un grupo de expertos en el lugar, lo que resulta en una mínima perturbación a las especies que lo habitan y costos operativos reducidos.

Realizar este tipo de análisis es un proceso que requiere horas e incluso días de trabajo si se realiza de forma manual, debido a que es necesario recolectar gran cantidad de grabaciones para luego identificar en cada una de estas, las vocalizaciones o llamados



de las especies de interés. Los algoritmos de detección automática de llamados y/o vocalizaciones de especies ofrecen una solución a este problema, empleando una variedad de técnicas de inteligencia computacional para detectar y clasificar vocalizaciones de animales de distintos grupos taxonómicos. Sin embargo, las diferentes propuestas presentan ventajas y desventajas a la hora de realizar el análisis de datos adquiridos por monitoreo acústico pasivo.

Actualmente, se cuenta con múltiples alternativas para la identificación automática de vocalizaciones y llamados de especies. Una de las alternativas presentadas es el software Avisoft [19] desarrollado por Avisoft Bioacoustics. Este software es una herramienta para el análisis de la comunicación acústica animal multi-especie incluyendo mamíferos, aves, roedores, ranas, peces e insectos [20] [21]. Esta herramienta implementa como método de clasificación un HMM (*Hidden Markov Models*) [22], el cual realiza un análisis detallado de las vocalizaciones por sílabas, y la manera en que se organizan para identificar cantos más complejos. Otro software presentado en literatura y ampliamente usado por expertos en bioacústica es Arbimon [6], desarrollado por Rainforest Connection. Este software utiliza como método de clasificación el algoritmo de Random Forest [12] y a diferencia de Avisoft, cuenta con una plataforma online para realizar el análisis de patrones [23], análisis de paisaje acústico [24], entrenamiento de modelos y reconocimiento en las bases de datos que se carguen en la nube [25]. Sin embargo, esta plataforma requiere que todas las grabaciones a analizar sean subidas a la nube dificultando el trabajo para grandes volúmenes de datos. Estos softwares, al igual que la mayoría, de métodos de identificación automática de llamados de especies requieren de etiquetas previas a realizar el entrenamiento de los modelos para las especies específicas [26].

Adicionalmente, se ha propuesto el uso de técnicas aprendizaje profundo para realizar la detección y clasificación de vocalizaciones, como es el caso de las Redes Neuronales Convolucionales (CNN) [9]. Las CNN se destacan por su buen desempeño con grandes conjuntos de datos, pero al igual que los métodos de aprendizaje de máquina usados en otras propuestas, se necesitan etiquetas precisas de las vocalizaciones objetivo para realizar el entrenamiento de la red, lo que limita su análisis a las especies previamente reconocidas por un experto. Estas redes convolucionales son usadas en la metodología propuesta por Bergler et al. 2022 ANIMAL-SPOT [3], esta metodología permite diferenciar vocalizaciones en presencia de ruido de fondo, identificación de animales de diferentes grupos taxonómicos y reconocimiento de múltiples tipos de llamadas.

En general, en estas propuestas se requiere un trabajo de etiquetado manual de las vocalizaciones en las grabaciones de audio, creación de plantillas, entrenamiento

especie por especie o modificación de parámetros tales como la banda de frecuencia a analizar o la ventana utilizada para generar los espectrogramas. De igual forma, los grupos SISTEMIC y GHA de la Universidad de Antioquia han trabajado de la mano en el desarrollo de una metodología que permita la identificación de especies de animales de manera no supervisada de tal forma que no se requieran etiquetas para entrenamiento del modelo y que permita hacer la detección de especies de las cuales no se tiene conocimiento a priori en la zona [16]. Esta metodología utiliza el algoritmo de clustering LAMDA [27] para agrupar las vocalizaciones de las especies en función de sus similitudes acústicas. La propuesta no supervisada presenta una ventaja en el análisis de datos en ecosistemas como los colombianos, que cuentan con una gran diversidad de especies entre las que se pueden encontrar algunas aún desconocidas para la ciencia. Además, la metodología propuesta no necesita entrenar un modelo especie-específico o cambiar parámetros específicos por especie. En este caso, el algoritmo propone sonotipos a los que, en una etapa posterior, un experto asociará con especies y que podrá reconocer en nuevas grabaciones de audio.

En este trabajo se hizo un análisis y comparación de los distintos algoritmos de aprendizaje (supervisado y no supervisado) propuestos para la detección automática de llamados y vocalizaciones de múltiples especies de animales presentes en paisajes sonoros. Se analizaron los desempeños de cada propuesta con una base de datos de audios adquiridos en zonas del bosque húmedo colombiano. El objetivo fue identificar las ventajas y desventajas que presenta cada alternativa, de tal manera que se pueda hacer una selección de la metodología a usar de acuerdo al problema que se quiera abordar.

### **Base de datos y sitio de estudio**

Para el desarrollo del proyecto se usaron dos conjuntos de grabaciones de audios obtenidas por el Grupo Herpetológico de Antioquia (que hace parte del Instituto de Biología de la Universidad de Antioquia). Estas bases de datos fueron adquiridas en Puerto Wilches, Santander, Colombia (7°21'52.5" N, 73°51'33.0" W), ilustrada en la figura 1. La zona analizada corresponde a un área circular delimitada en un radio de 1500m en la que predominan cultivos de palmas de aceite (75%), vegetación secundaria (7.6%), bosque (6.13%), pastizal (5.5%) y vegetación acuática (3.2%).

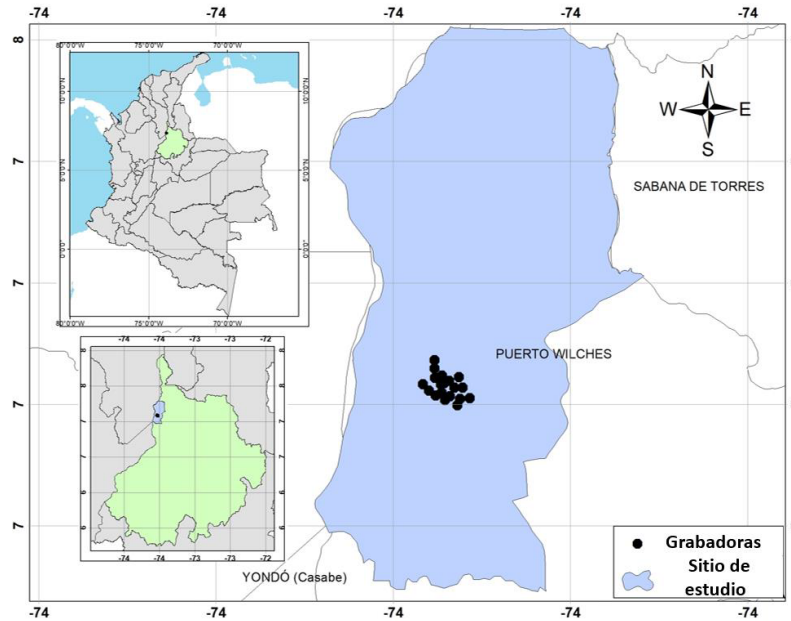


Figura 1. Mapa de la ubicación del sitio de estudio. Tomado de Restrepo-Rivas, 2022 [28]

El primer *dataset* consta de 207 grabaciones extraídas de un total de 2638 grabaciones de audio en las que, luego de un estricto trabajo de etiquetado por parte de expertos, se encontraron 11 especies, incluyendo seis especies de aves, 4 especies de anuros y un primate. Este *dataset* fue recolectado utilizando grabadoras Song Meter Mini, configuradas para grabar un minuto cada 10 minutos con una tasa de muestreo de 48kHz. El segundo *dataset* consta de 197 grabaciones de audio y está enfocado en especies en el espectro ultrasónico. Para la recolección de estos datos, se usó grabadoras Song Meter Mini bat para grabar 15 segundos cada 15 minutos con una tasa de muestreo de 384kHz. En esta base de datos se encontraron 13 especies de murciélagos y 6 ortópteros.

## Metodología

En esta sección se describe la metodología propuesta para el desarrollo del trabajo. En la figura 2 se muestran las 4 etapas principales que se llevaron a cabo en este trabajo: revisión del estado del arte, selección de algoritmos, pruebas con los algoritmos seleccionados y finalmente una evaluación y comparación de sus desempeños.

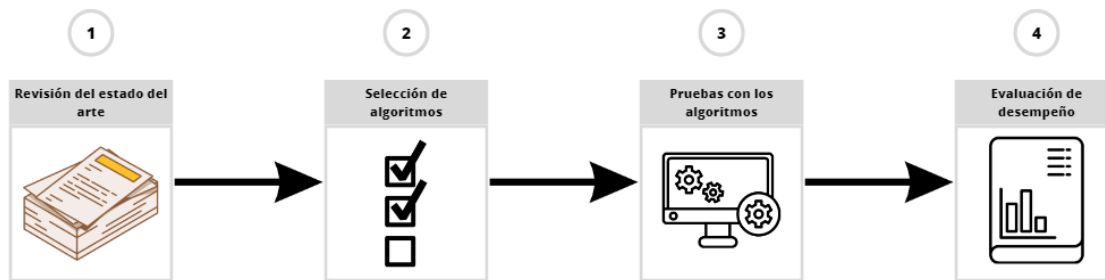


Figura 2. Metodología propuesta.

### Revisión del estado del arte

Se realizó una búsqueda bibliográfica enfocada en los *softwares* y metodologías utilizadas para la detección y creación de modelos de clasificación de especies de distintos grupos taxonómicos en datos de monitoreo acústico pasivo. Para este proceso, se consultaron fuentes destacadas como Scencedirect, Google Scholar, BES Journals y revistas especializadas en ciencias biológicas, como PeerJ - Life and environment y Methods in Ecology and Evolution. Durante esta búsqueda, se utilizaron términos y palabras clave como: “*Ecoacoustics*”, “*Automatic detection*”, “*Passive acoustic monitoring*”, “*Classification*” y “*Software*”. Posteriormente, se procedió a la lectura y análisis de los documentos encontrados. Aquellos artículos considerados relevantes por la propuesta de *software* o metodologías de detección automática de llamados de animales, en particular aquellos que incluían múltiples especies en sus aplicaciones, fueron organizados en una tabla. Esta tabla presenta el nombre del artículo, el uso específico del software o metodología, los grupos taxonómicos analizados y la existencia de una interfaz gráfica de usuario (ver anexo 1). Un ejemplo de esta recopilación se puede observar en la Tabla 1.

Nombre	Uso	Grupos Taxonómicos	GUI
Real-time bioacoustics monitoring and automated species identification	Creación de modelos especie-específicos para anfibios, aves, mamíferos e insectos	Anfibios, aves, mamíferos e insectos	Si
A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network	Utilizaron Arbimon para el almacenamiento, visualización y creación de una base de datos de entrenamiento utilizando la herramienta de pattern matching	Aves, anfibios	No
Climate change is creating a mismatch between protected areas and suitable habitats for frogs and birds in Puerto	Se utilizó arbimon para almacenar, hacer pattern matching y analizar presencia y	Aves, anfibios	Si

Rico	ausencia, verificada luego manualmente		
Occupancy and Abundance of Eleutherodactylus Frogs in Coffee Plantations in Puerto Rico	Almacenamiento y análisis de los datos con Arbimon y random forest	Anfibios	Si
Applications and advances in acoustic monitoring for infectious disease epidemiology	Análisis de frecuencia de detección de especies en los sitios de investigación	Anfibios, insectos, aves y primates	Si
Passive Acoustic Monitoring as a Tool to Investigate the Spatial Distribution of Invasive Alien Species	Se almacenaron los datos para hacer un modelo de ocupación de especies invasoras	Anfibios, aves y mamíferos	Si
Software performance for the automated identification of bird vocalizations: the case of two closely related species	Band limited energy detector para la detección de eventos de audio	Aves	Si
PNW-Cnet v4: Automated Species Identification for Passive Acoustic Monitoring	Se creó un sistema para la detección de múltiples especies	Aves, anfibios, mamíferos e insectos	Si
Automatic Detection and Compression for Passive Acoustic Monitoring of the African Forest Elephant	Se creó un sistema para el monitoreo de Elefante africano	Elefantes	No

Tabla 1. Ejemplo de recopilación de bibliografía.

### Selección de algoritmos

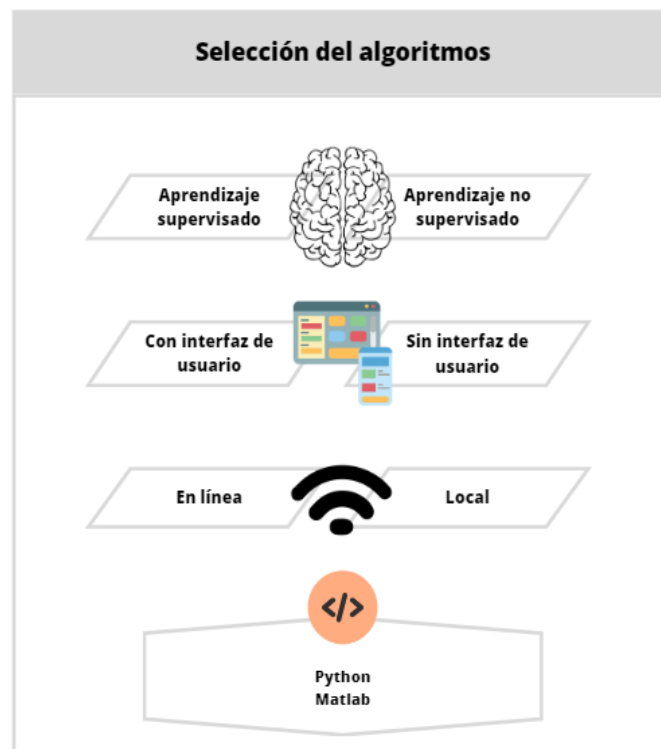


Figura 3. Criterios de selección de algoritmos.

A partir de los artículos analizados en la etapa de revisión del estado del arte, se seleccionaron diversos *softwares* y metodologías, con el fin de representar la variedad de criterios presentados en la figura 3. Esto significa que se optó por software con interfaz gráfica de usuario y sin ella, se incluyeron metodologías basadas en aprendizaje supervisado y no supervisado. Se consideraron opciones de *software* con servicios web y almacenamiento en la nube, así como las que se basan exclusivamente en archivos locales. Además, se incluyeron diferentes lenguajes de programación, como MATLAB y Python. El propósito de esta selección diversa fue abarcar la mayor variedad posible de implementaciones de *software* y metodologías para la detección automática de animales de diferentes grupos taxonómicos, desde el lenguaje de programación utilizado en el algoritmo, hasta la experiencia del usuario.

### ***Pruebas con los algoritmos***

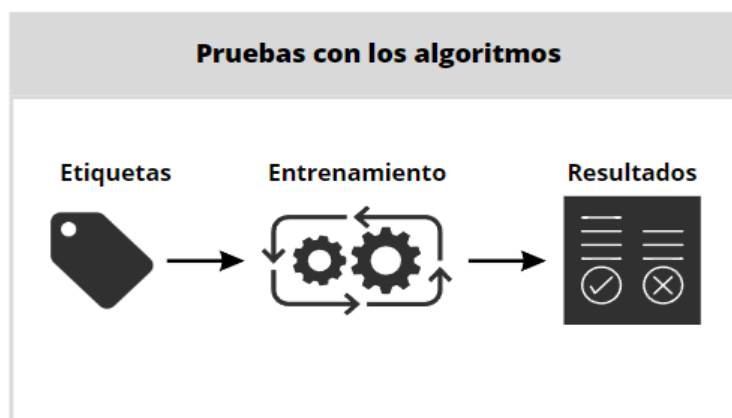


Figura 4. Metodología para pruebas con los algoritmos seleccionados.

En general, las pruebas realizadas con cada algoritmo se desglosan en tres etapas, tal como se ilustra en la figura 4. En primer lugar, se asignan etiquetas a los llamados presentes en la base de datos. Estas etiquetas proporcionan información importante tales como frecuencias mínima y máxima, comienzo y final de la vocalización, el audio en el que se encuentra y una clase asignada. Es relevante destacar que algunos softwares y metodologías ofrecen herramientas que facilitan esta labor, adaptando las etiquetas al formato requerido para las etapas subsiguientes.

En la siguiente etapa, una vez etiquetada la base de datos, se procede al entrenamiento de los algoritmos. En este proceso, se ajustan los parámetros de cada algoritmo para adaptarse a los diferentes tipos de llamados. Por lo general estos parámetros constan de

la banda de frecuencia, un intervalo de tiempo que determina la duración mínima y máxima de la vocalización y un número de épocas de entrenamiento.

Finalmente, En la última etapa, se recopilan y analizan los resultados proporcionados por cada metodología. Algunos softwares ofrecen herramientas que generan estos resultados automáticamente. Sin embargo, en aquellos casos donde no es así, los resultados se determinan de forma manual, comparando las detecciones realizadas por el software con las etiquetas previamente establecidas en la base de datos.

### ***Evaluación de desempeño***

Para evaluar el desempeño de los algoritmos, se utilizaron las matrices de confusión, las cuales se construyeron tras comparar las etiquetas originales de la base de datos con las predicciones generadas por cada modelo. Estas matrices proporcionan los siguientes valores:

- TP (positivos verdaderos): corresponde a las presencias que fueron correctamente identificadas.
- TN (negativos verdaderos): representa las ausencias que fueron correctamente identificadas.
- FP (falsos positivos): denota las presencias que fueron incorrectamente clasificadas como ausencias.
- FN (falsos negativos): refiere a las ausencias que fueron incorrectamente clasificadas como presencias.

### ***Métricas de evaluación***

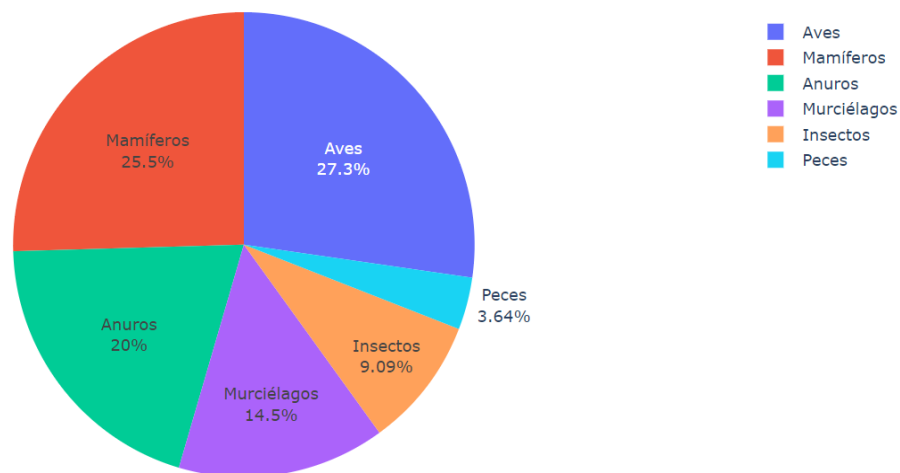
La métrica F1-Score es ampliamente utilizada en el campo de la inteligencia computacional para calcular la exactitud de un modelo cuando la base de datos no está balanceada para todas las clases a analizar. Esta métrica se calcula a partir de la matriz de confusión generada a partir de las detecciones para cada una de las metodologías analizadas usando la ecuación (1).

$$F1 - Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (1)$$

## Resultados y análisis

### *Revisión sistemática del estado del arte*

En la revisión bibliográfica realizada, siguiendo la metodología descrita en la sección revisión del estado del arte, se recopiló un total de 40 artículos con variedad de propuestas de algoritmos y diferentes grupos taxonómicos trabajados, distribuidos como se muestra en la figura 5. En esta figura se observa que, de la literatura revisada, las aves son el grupo taxonómico más estudiado, representando el 27.3% del total de artículos revisados, seguido por los mamíferos con un 25.5% y los anfibios con un 20%.



*Figura 5. Distribución de los artículos por grupo taxonómico que detectan con el método propuesto en el artículo.*

La figura 6 ilustra la distribución de los artículos en función de los conjuntos de grupos taxonómicos analizados en las metodologías propuestas. En esta figura resalta la prevalencia del estudio de aves y anfibios en conjunto con un 40%, superando a los demás grupos.



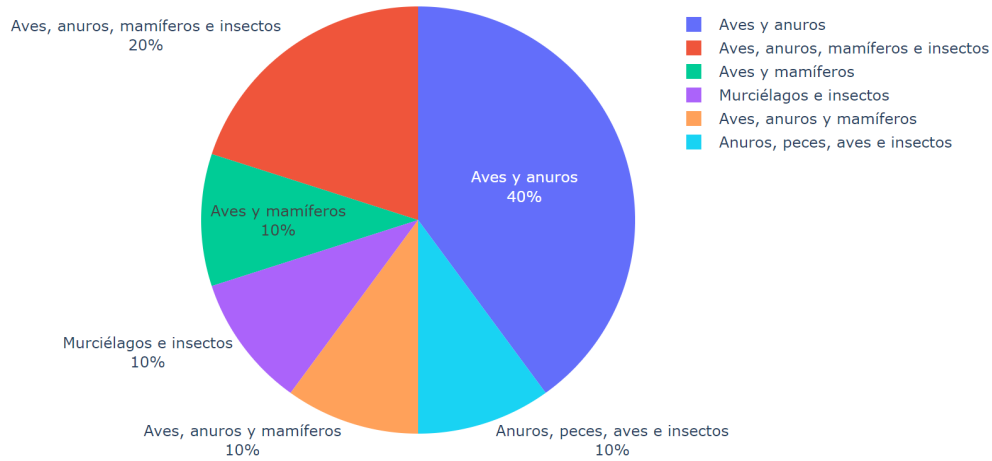


Figura 6. Distribución de los artículos por conjunto de grupos taxonómicos trabajados.

Adicionalmente, se realizó un análisis de la implementación de interfaz gráfica de usuario (GUI) en las metodologías identificadas en el estado del arte. El objetivo era determinar si la interfaz de usuario es relevante en el uso de softwares para reconocimiento automático de llamadas de especies. Los resultados del análisis revelan una marcada inclinación hacia la implementación de GUI en los softwares y metodologías, sobre las soluciones que funcionan exclusivamente mediante código. Esto se evidencia en la figura 7, con un 79.1% de artículos que muestran softwares y metodologías con GUI implementada, mientras que solo un 20.9% carecen de esta.

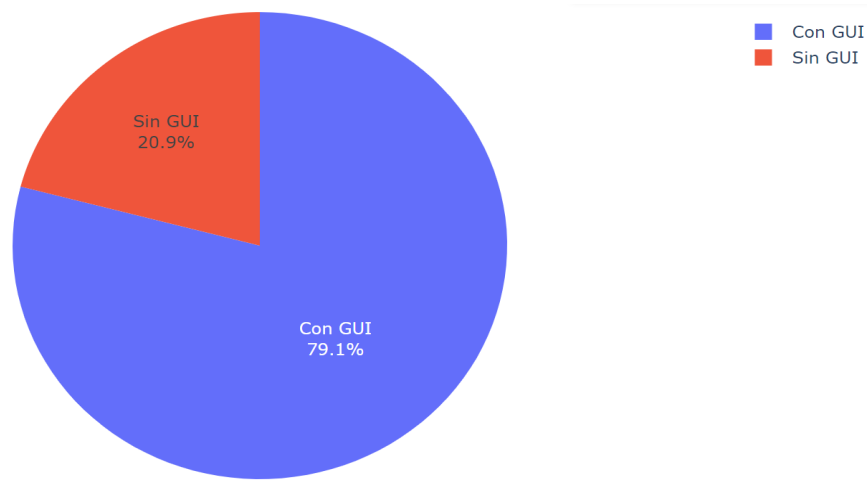


Figura 7. Distribución de los artículos que trabajaron con o sin interfaz de usuario.

### **Selección de software y metodologías**

Siguiendo los criterios mencionados anteriormente en la sección revisión del estado del arte, se seleccionaron 4 propuestas de softwares y 1 metodología para realizar la

identificación de las vocalizaciones/llamados de la base de datos propuesta y evaluar el desempeño para cada propuesta. La selección realizada fue la siguiente:

Kaleidoscope Pro [29]: Es un software cuyo uso requiere una licencia de pago. Este software cuenta con interfaz de usuario, equipada con herramientas para análisis de *clusters* utilizando un algoritmo de modelo oculto de Markov (HMM), *spectral clustering* y realiza almacenamiento de audios en la nube. A diferencia de los softwares descritos anteriormente, Kaleidoscope Pro cuenta con un módulo específico para el análisis de audio en el espectro ultrasónico diseñada especialmente para el análisis y reconocimiento de murciélagos.

Arbimon [6]: Es una interfaz web gratuita diseñada para almacenar en la nube y analizar grandes conjuntos de datos de monitoreo acústico pasivo. Tiene una interfaz de usuario intuitiva y cuenta con gran variedad de herramientas para tareas de clasificación de audios en listas de reproducción y sitios de grabación, análisis de audio con reconocimiento de patrones y creación de modelos de clasificación de vocalizaciones utilizando un algoritmo de *Random Forest (RF)*. Arbimon resulta relevante para la elección de software debido a su plataforma en línea, almacenamiento en la nube y a la prominencia que tiene en el estado del arte. Numerosos proyectos recurren a esta plataforma no solo para el almacenamiento y etiquetado de bases de datos, sino también para la elaboración de modelos de clasificación de animales de diversos grupos taxonómicos. Adicionalmente, se está desarrollando una herramienta de *clustering* dentro de esta plataforma para la detección y clasificación de vocalizaciones. Esta herramienta se tuvo en cuenta también para el análisis.

Aureas [16]: es una metodología no supervisada desarrollada por los grupos de investigación SISTEMIC – GHA de la Universidad de Antioquia. Esta propuesta implementa el algoritmo LAMDA 3pi (*Learning Algorithm for Multivariate Data Analysis*) para el análisis de datos acústicos sin necesidad de definir etiquetas previas en la base de datos. La metodología permite el análisis tanto en el espectro audible como el ultrasonido para múltiples especies de forma simultánea. Cuenta con una etapa de segmentación que permite aislar los posibles eventos acústicos o llamados de especies, seguida por una etapa de extracción de características en la que extrae la información acústica relevante a cada segmento. Finalmente entra a una etapa de clustering que agrupa los segmentos de acuerdo a la similitud acústica presentada, permitiendo encontrar de manera simultánea, múltiples especies asociadas a diferentes grupos taxonómicos. A diferencia de los algoritmos anteriormente mencionados, no requiere la definición de parámetros por parte del usuario.

Raven Pro / Koogu [30]: al igual que Kaleidoscope pro, destaca en este estudio por su prominente posición en el estado del arte. Este software es comúnmente utilizado para la visualización de espectrogramas y etiquetado manual de audios. Sin embargo, recientemente se incorporó una funcionalidad que permite integrar un modelo de reconocimiento automático generado mediante el paquete de Python Koogu. Esta herramienta utiliza los audios y las tablas de selección proporcionados por Raven Pro para entrenar un modelo basado en redes neuronales convolucionales.

ANIMAL-SPOT [3]: es una alternativa seleccionada para representar una metodología que utiliza deep learning. Esta metodología implementa una red neuronal convolucional (CNN) basada en la arquitectura ResNet-18 (red convolucional con 18 capas de profundidad), desarrollada en Python sin interfaz gráfica de usuario. Esta alternativa cuenta con herramientas para la detección binaria que permite diferenciar una llamada objetivo del ruido de fondo, identificación de múltiples clases (especies) y reconocimiento de múltiples tipos de llamadas.

### ***Descripción de parámetros y pruebas con las metodologías seleccionadas***

#### ***- Kaleidoscope Pro:***

A simple vista, el software presenta una interfaz gráfica intuitiva y de diseño atractivo. Aunque su requerimiento inicial es la base de datos de grabaciones de audio que contienen las especies de interés, ofrece una funcionalidad adicional de importar tablas de detección previamente generadas. Esto permite reproducir resultados de experimentos previos o realizar nuevas detecciones basadas en dichas tablas. No obstante, la configuración requerida para iniciar el proceso de clustering implica tener un conocimiento previo sobre el significado y efecto de cada uno de los parámetros mostrados enunciados a continuación y representados en la figura 8.

CLUSTER ANALYSIS

Scan and cluster recordings to create cluster.kcs and cluster.csv

1.0 Max distance from cluster center to include outputs in cluster.csv (2.0 max)

FFT Window 5.33ms (128 @0-12kHz 256 @13-24kHz 512 @25-48kHz 1024 @49-96kHz)

12 Max states (8-16 is a good range)

.5 Max distance to cluster center for building clusters (0.5 is recommended)

500 Max clusters (default 500)

Figura 8. Selección de parámetros para análisis de clusters en Kaleidoscope Pro.

*Max distance* (Distancia máxima): Este parámetro se refiere a la distancia máxima desde el centro del cluster para incluir las salidas en el archivo de resultados cluster.csv. Funciona como una especie de filtro que excluye del análisis aquellas señales cuya distancia al centro del cluster supere el valor especificado.

*FFT Window size* (Tamaño de la ventana FFT): Se utiliza para influir en el análisis espectral de los archivos de audio de entrada, y se adapta a las características específicas de las grabaciones.

*Max States* (Máximo de Estados): Cambia el tamaño del objetivo del modelo oculto de Markov. Un mayor número de estados puede ser necesario en ambientes con mayor diversidad de señales acústicas y puede ayudar a distinguir diferencias sutiles entre vocalizaciones similares. Sin embargo, un valor demasiado grande puede generar resultados ruidosos y pobres en el análisis de *clustering*.

*Max Distance to Cluster Center for Building Clusters* (Distancia Máxima al Centro del Cluster para la Formación de Clusters): Este parámetro controla cómo las señales detectadas se agrupan en *clusters*. Si la señal detectada está dentro de la distancia especificada con respecto a un *cluster* existente, se asigna a este *cluster*. De lo contrario, se puede formar un nuevo cluster para incluir la señal detectada. Esta distancia también se utiliza para fusionar dos *clusters*.

*Max Clusters* (Máximo de Clusters): Este parámetro establece el número máximo de clusters que pueden formarse. Los clusters que contienen un alto número de señales detectadas similares tienen prioridad, mientras que los clusters con señales menos similares se descartan para ajustarse a este valor máximo.

La configuración de estos parámetros plantea principalmente 2 desafíos. En primer lugar, presenta un obstáculo para un experto en biología sin conocimiento específico de la técnica de clustering implementada y el significado de los parámetros que deben ajustarse. En segundo lugar, estas configuraciones deben realizarse de manera individual para cada especie que se desee analizar, lo que convierte el proceso en una tarea más laboriosa y menos eficiente cuando se trabaja con múltiples especies.

Los resultados del análisis se muestran en dos ventanas (Figura 9). En la primera ventana se tiene una tabla con información de cada uno de los segmentos obtenidos, mientras que en la segunda se visualiza la sección del espectrograma correspondiente donde además se puede reproducir el audio y trasladarse entre audios y segmentos de manera fluida.

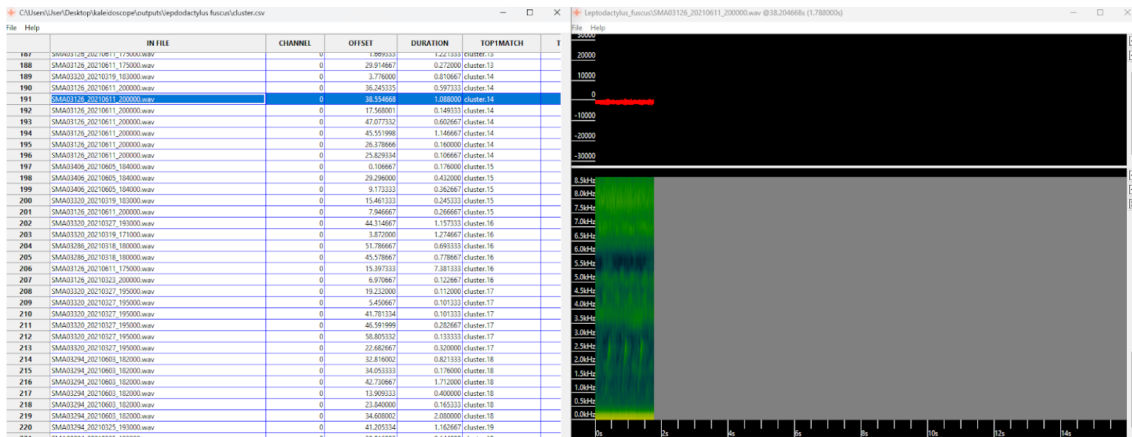


Figura 9. Ventanas de resultados en Kaleidoscope.

Sin embargo, los resultados encontrados con este software no presentaron el desempeño esperado. En varios de los casos analizados, se observaron segmentos de larga duración que, en vez de focalizarse en una única vocalización, capturaban una diversidad de cantos. Estos segmentos se clasificaban según el patrón de mayor intensidad del espectrograma presente. Como resultado, muchos de los clusters identificados correspondían a estridulaciones de insectos con frecuencias consistentes y de alta intensidad. Además, había algunos clusters que contenían una combinación de vocalizaciones de diversas especies, lo que complicaba la tarea de seleccionar un cluster representativo para una especie en particular. La figura 10 presenta los resultados del análisis de clusters realizado por Kaleidoscope Pro para 11 especies en el espectro audible y 7 murciélagos en el espectro ultrasónico de la base de datos analizada evaluando su desempeño con la métrica F1-Score.

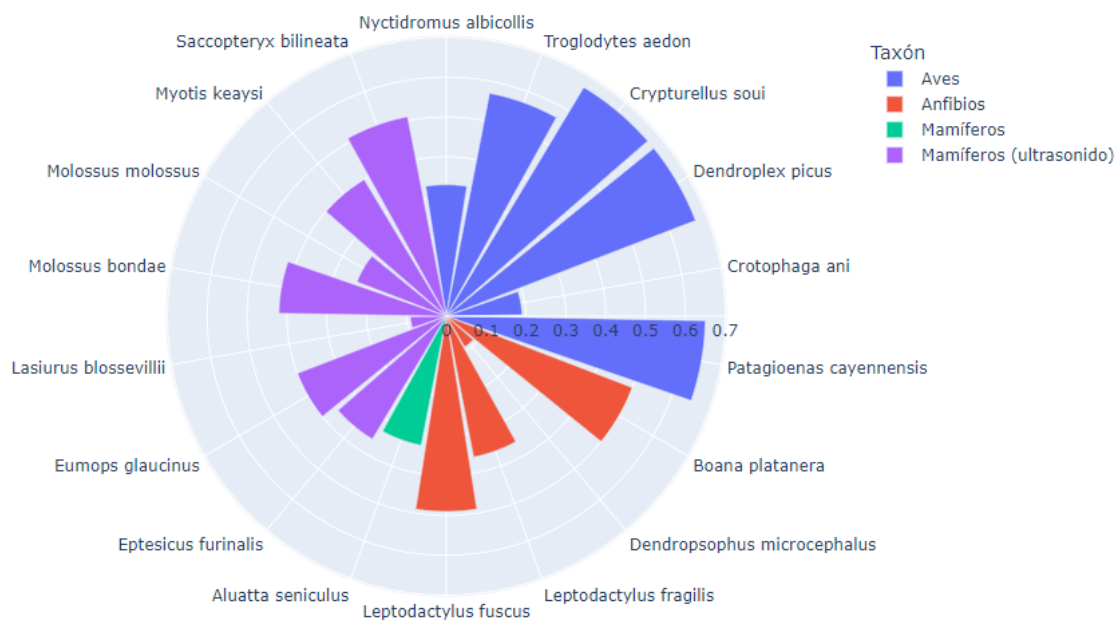


Figura 10. Resultados F1 score kaleidoscope Pro

En los resultados se hace evidente el efecto de la agrupación de variedad de cantos en un solo segmento. En el caso de los anfibios, cuyas vocalizaciones son estereotipadas y repetitivas, comparten el paisaje sonoro con grillos cuyas estridulaciones aparecen más intensas en el espectrograma. Esto causó confusión en la etapa de segmentación debido a que el software tomó en un solo segmento, varios llamados de una misma especie, generando segmentos más grandes que fueron agrupados en clusters de insectos. En cuanto al mamífero y los murciélagos, la intensidad de sus vocalizaciones en el espectrograma no fue lo suficientemente destacada como para generar segmentos adecuados que permitieran la identificación de la especie en un cluster, lo que resultó en un rendimiento notablemente bajo en su identificación. Por otro lado, en el caso de aquellas aves que presentan una mayor cantidad de vocalizaciones con una alta intensidad en el espectrograma, obtuvieron un rendimiento generalmente superior en comparación con los tres grupos taxonómicos anteriores.

- ***Arbimon Random Forest:***

El primer paso para utilizar esta herramienta consiste en cargar la base de datos de audios a la plataforma web de Arbimon, y organizarlos en forma de listas de reproducción. Estas listas de reproducción se utilizan posteriormente para llevar a cabo el entrenamiento del modelo de identificación automática. La herramienta de clasificación de audios mediante listas de reproducción demostró ser altamente útil durante el proceso de etiquetado de la base de datos en la plataforma web. La utilidad de la herramienta de reconocimiento de patrones facilita la tarea de etiquetar los audios que contienen la especie que se está identificando. Asimismo, facilita la definición del patrón de búsqueda necesario para el posterior entrenamiento del modelo de *random forest*. Para proceder con este entrenamiento, solo se necesitan las etiquetas de presencia y ausencia previamente asignadas, definir una región de interés y la selección de una partición de los audios etiquetados destinada al entrenamiento y validación del modelo, tal como se ilustra en la figura 11.

### Create a new model

**Model name**  
Model\_Patagioenas

**Classifier** Classifiers may change without notice. More information.  
Pattern Matching with Random Forests

**Training Set**  
Patagioenas cayennensis

**Validated recordings available** ⓘ  
for: Patagioenas cayennensis (Common Song)

	Total: 44	Use in fitting:	Use in validation:
Present: 22		16	6 /6
Absent: 22		16	6 /6

Create Close

Figura 11. Configuración de modelo de random forest en arbimon

Tras entrenar los modelos para cada especie dentro de la base de datos del espectro audible, se obtuvieron los resultados plasmados en la Figura 12. En términos generales, se alcanzó un F1-Score cercano a 0.7. Sin embargo, cabe destacar que para las especies *Leptodactylus fragilis* y *Leptodactylus fuscus* no se logró identificar correctamente ninguna presencia, lo que condujo a un F1-Score de 0 para ambas. Por otro lado, el ave *Crotophaga ani* presentó un desempeño inferior, con detecciones muy escasas, lo que resultó en un F1-Score de 0.4.

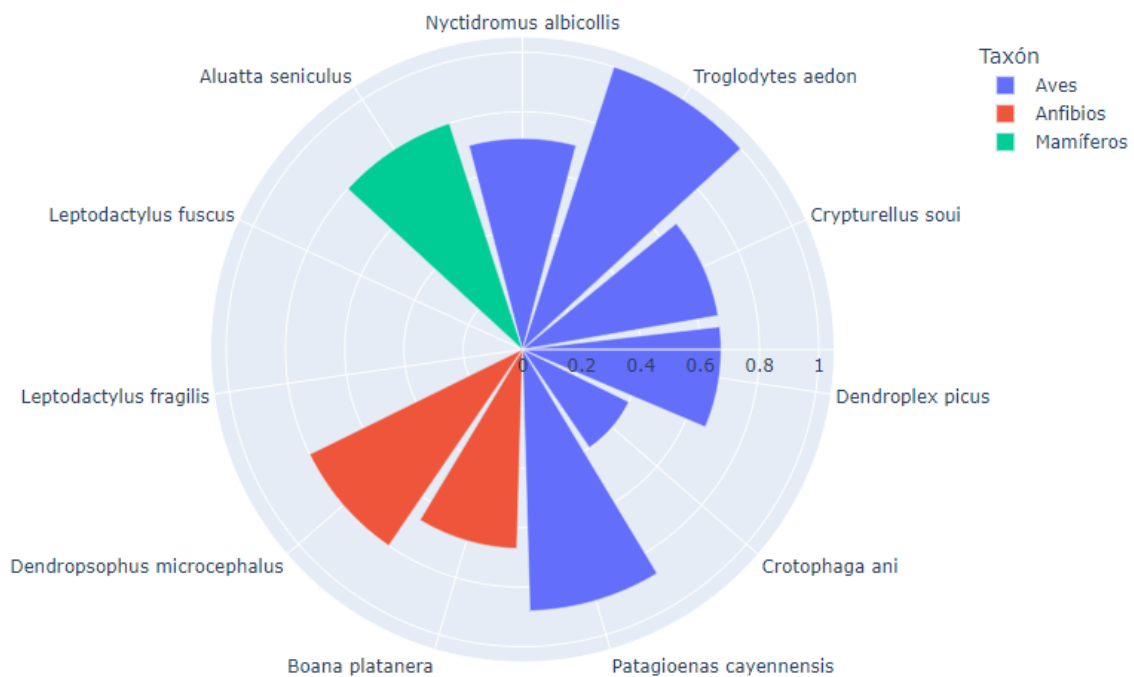


Figura 12. Resultados de F1 score Arbimon con clasificador Random Forest

- **Arbimon Clustering (BETA):**

Arbimon se encuentra desarrollando una herramienta de clustering con 2 aplicaciones por el momento: La detección de eventos acústicos (AED) y el análisis de clusters basados en los eventos acústicos detectados. Esta herramienta emplea para el análisis la misma base de datos de audios cargada y etiquetada que se utiliza para entrenar el modelo de *random forest*. Para la detección de eventos acústicos se seleccionan parámetros básicos como lo son una frecuencia mínima y máxima, duración del canto, entre otros como se muestra en la figura 13. Es importante destacar que los parámetros *duration threshold*, *bandwidth threshold* y *area threshold* son especialmente sensibles, y determinan en conjunto la efectividad de la detección de eventos acústicos. Estos parámetros otorgan a la herramienta la capacidad de adaptarse a las características específicas de las vocalizaciones de cada especie, incluyendo su duración y el ancho de banda que ocupa en el espectrograma. Una configuración deficiente de estos parámetros puede resultar en una alta cantidad de eventos acústicos irrelevantes o en que el detector no logre identificar suficientes segmentos correspondientes a vocalizaciones de una especie, lo que puede tener repercusiones en el proceso de clustering posterior.

Amplitude Threshold=1, Duration Threshold=0.1, Bandwidth Threshold=1, Area Threshold=0.2, Filter Size=10, Min Frequency=1.5, Max Frequency=7	SANTIAGO TABORDA DIOSA	100
--	------------------------	-----

Figura 13. Configuración AED arbimon clustering

Luego, para realizar el análisis de clusters, se configuran parámetros que determinarán la cantidad y tamaño de los clusters generados por el algoritmo de Arbimon como se muestra en las figuras 14 y 15.



### Create a new clustering job

**Job Name**

**Audio Event Detection Job**

**Distance Threshold** ⓘ

**Min. Points** ⓘ

**Max. Cluster Size** ⓘ

Figura 14. Configuración de análisis de clusters en arbimon

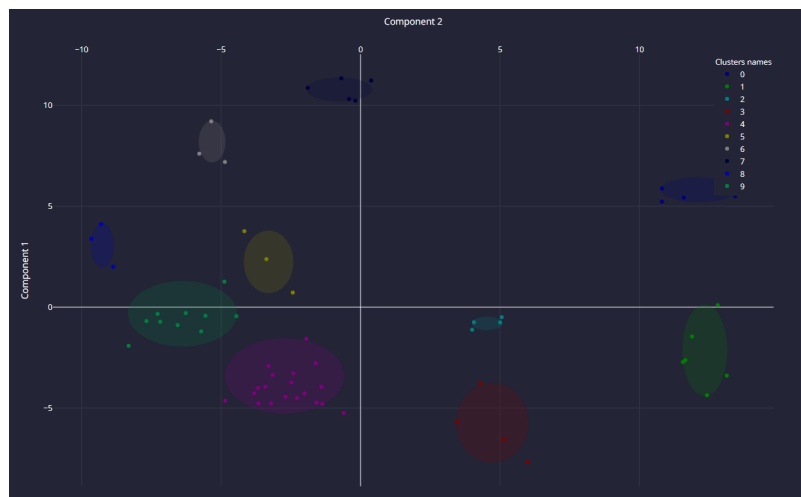


Figura 15. Clusters generados con la configuración mostrada en la figura 14.

Obtener resultados para esta metodología resultó complejo debido a que aún no se ha implementado una herramienta que permita comparar las detecciones obtenidas con las etiquetas de la base de datos. A pesar de esto, el análisis de clusters (presentado en la figura 16) mostró resultados similares a los del modelo de *random forest*. Al igual que en la metodología previa, no se lograron detecciones para las dos especies de anfibios que ya habían presentado dificultades. Además, se observó un rendimiento reducido en las otras dos especies de anfibios. Por otro lado, es destacable que el desempeño en la identificación de mamíferos y aves experimentó una mejora considerable.

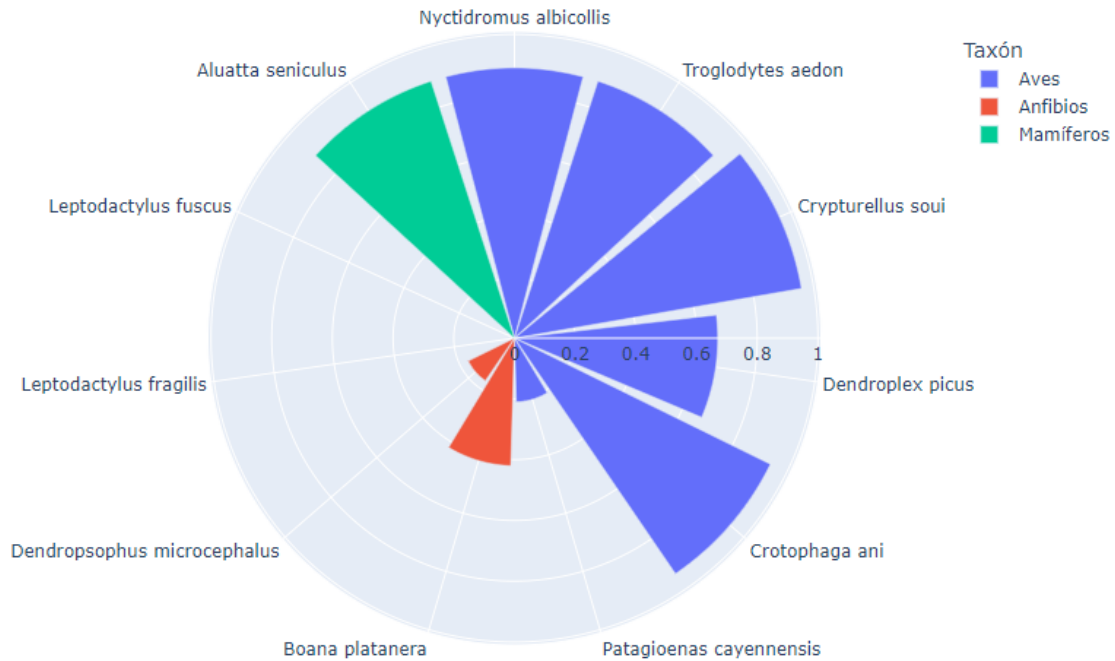


Figura 16. Resultados F1 score arbimon clustering

- **Aureas:**

Al igual que en Kaleidoscope, Aureas toma como información de entrada una base de datos de audio de manera local y ofrece la posibilidad de cargar información de detecciones anteriores. La interfaz de Aureas, desarrollada en MATLAB, es intuitiva y facilita el proceso de selección de audios para la segmentación automática y generación de clusters. Al culminar estas etapas, se presenta una tabla con detalles sobre el audio correspondiente, así como características de los segmentos relacionados con rangos de frecuencias y tiempo. No obstante, la clasificación de los segmentos en clusters puede presentar ciertas dificultades, ya que los segmentos no se muestran directamente y deben ser seleccionados para visualizar la porción del espectrograma que les corresponde. Para bases de datos extensas, este proceso puede volverse más lento a medida que se avanza a través de los segmentos. A pesar de esto, una fortaleza de Aureas es la capacidad de definir en uno o varios clusters, el llamado de las especies de interés. Adicionalmente, ofrece la opción de "limpiar" los clusters, filtrando segmentos que no se alineen con la clase que se está buscando. Esto permite recalibrar las propiedades del cluster, facilitando su uso en etapas posteriores de reconocimiento.

Una de las ventajas a destacar del modelo de clustering difuso de AUREAS es la obtención del grado de pertenencia de los segmentos al cluster al que han sido asignados. Con esta característica es posible llevar a cabo un procesamiento posterior a los datos, estableciendo un umbral en el grado de pertenencia de los segmentos. Este

umbral permite reducir la cantidad de falsos positivos admitidos en las detecciones finales, mejorando así la precisión y fiabilidad de los resultados.

Los resultados obtenidos con Aureas (Figura 17) presentan un desempeño alto con valores de la métrica F1-Score por encima de 0.7 para todos los grupos taxonómicos, incluidos los anfibios, la principal deficiencia en las propuestas trabajadas anteriormente.

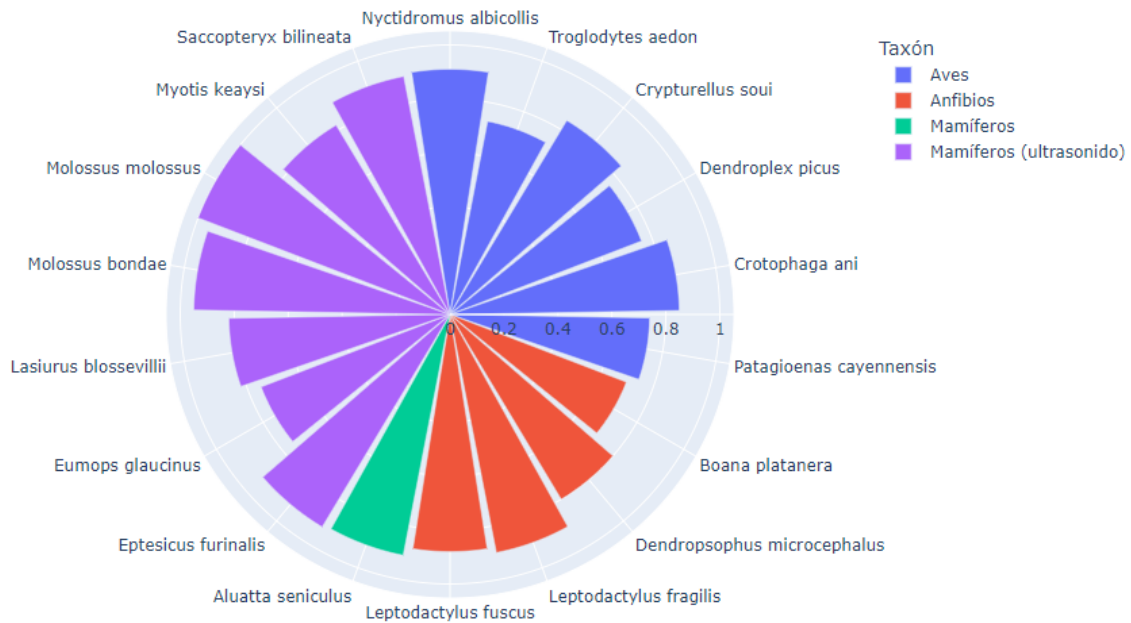


Figura 17. Resultados F1 Score Aureas

**- Koogu / Raven Pro:**

Raven Pro es reconocido y ampliamente utilizado debido a sus herramientas dedicadas a la visualización de espectrogramas y al etiquetado manual de especies en audios. Durante el desarrollo de este estudio, se pudo comprobar la eficacia y practicidad de este software para tales propósitos, ya que facilita la creación y exportación de tablas etiquetadas por clase de manera rápida y sencilla.

Dado que Koogu está diseñado para reconocer las etiquetas de una tabla de selección de Raven Pro, iniciar el entrenamiento del modelo de red neuronal es un proceso directo. Simplemente se requiere colocar las etiquetas y los audios en las carpetas adecuadas, y configurar parámetros esenciales como el ancho de banda a analizar y las épocas de entrenamiento. Sin embargo, durante este proceso, se detectó un comportamiento inusual: un error se presentaba en el código al definir un ancho de banda demasiado próximo a los valores mínimo y máximo de los segmentos etiquetados en Raven Pro. Por esta razón, se optó por utilizar anchos de banda más amplios. Una vez finalizado el

entrenamiento del modelo de red neuronal convolucional en Koogu, para ponerlo en funcionamiento en Raven Pro, basta con tomar los archivos generados por Koogu e incorporarlos en la carpeta de modelos de Raven, junto con dos archivos CSV que contienen las clases a identificar y sus respectivas etiquetas.

Todo este proceso requiere un conocimiento básico en Python, por lo que puede ser una alternativa menos llamativa o práctica para los expertos en biología, quienes son los más interesados en este tipo de softwares.

Los resultados obtenidos con esta metodología fueron los mejores junto con los de Aureas, siendo muy consistentes para todas las especies de los diferentes grupos taxonómicos como se muestra en la figura 18 a continuación.

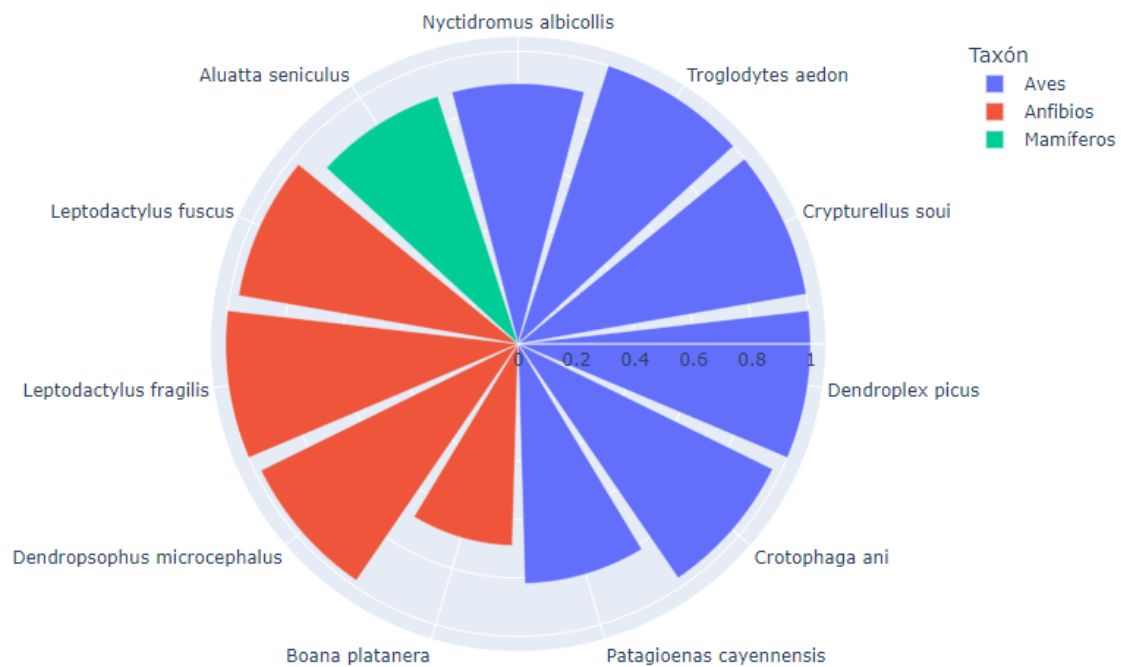


Figura 18. Resultados F1 Score Koogu

#### - ANIMAL-SPOT:

La implementación de ANIMAL-SPOT inició con la generación de etiquetas manuales para la base de datos de audios, lo que implicó cortar los segmentos de audio y asignarles un nombre con un formato que incluye información sobre el segmento de audio y la clase a la que pertenece. Una vez completado este proceso de etiquetado, se utilizó una herramienta que incorpora ANIMAL-SPOT para dividir los segmentos en carpetas de entrenamiento, prueba y validación. En este punto se encontró un desafío importante ya que la partición automática realizada por la herramienta no aseguraba la presencia de las especies en las 3 carpetas, dando lugar a errores. Incluso, cuando la

partición si conseguía incluir a cada especie en las 3 carpetas, esta resultaba ser desbalanceada en términos de la distribución de las clases. Para abordar este problema, se desarrolló un código adicional con el fin de reorganizar las carpetas y lograr una distribución equilibrada de las clases.

Durante el proceso de entrenamiento de los modelos, se enfrentaron desafíos adicionales. Inicialmente, se encontró una incompatibilidad con el sistema operativo Windows, ya que el algoritmo estaba diseñado para ejecutarse en Linux. Este inconveniente se resolvió mediante la modificación de una variable relacionada con el número de núcleos utilizados para los cálculos, lo que permitió su ejecución.

Una vez superados los obstáculos técnicos, los resultados obtenidos en el entrenamiento de los modelos no cumplieron con las expectativas. Los desempeños fueron consistentemente bajos (con valores de precisión por debajo del 50%) y similares en todos los casos, lo que sugiere que los modelos no estaban entrenando de manera efectiva. Al realizar el reconocimiento, las detecciones se centraron en la especie *Leptodactylus fragilis* cuando se consideraron segmentos de 0.3 segundos de duración, y para segmentos de mayor duración (hasta 3 segundos), las detecciones se clasificaron casi en su totalidad como *Patagioenas cayennensis* o *Crypturellus soui*. Si bien este trabajo proporciona valiosas lecciones y una idea de qué esperar al utilizar esta metodología en futuras ocasiones, los resultados obtenidos no fueron considerados en la compilación de resultados finales.

La figura 19 presenta una síntesis de los resultados en el espectro audible para todas las metodologías evaluadas en este estudio. En el eje horizontal se presentan los distintos softwares y metodologías, mientras que en el eje vertical se muestra el valor del F1-Score. Los bloques de colores indican el valor promedio del F1-Score para los diferentes grupos taxonómicos en cada propuesta. Es notable cómo Aureas y Koogu/Raven Pro sobresalen con los desempeños más altos, superando claramente a las metodologías de Arbimon y situando a Kaleidoscope Pro en la posición final.

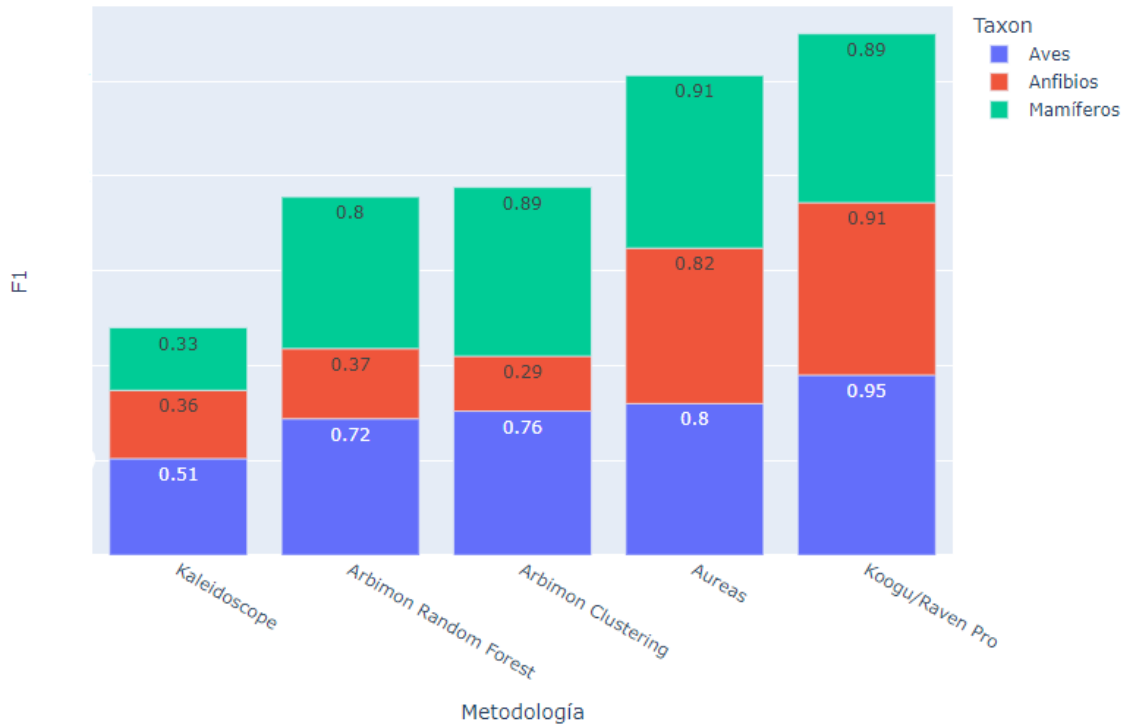
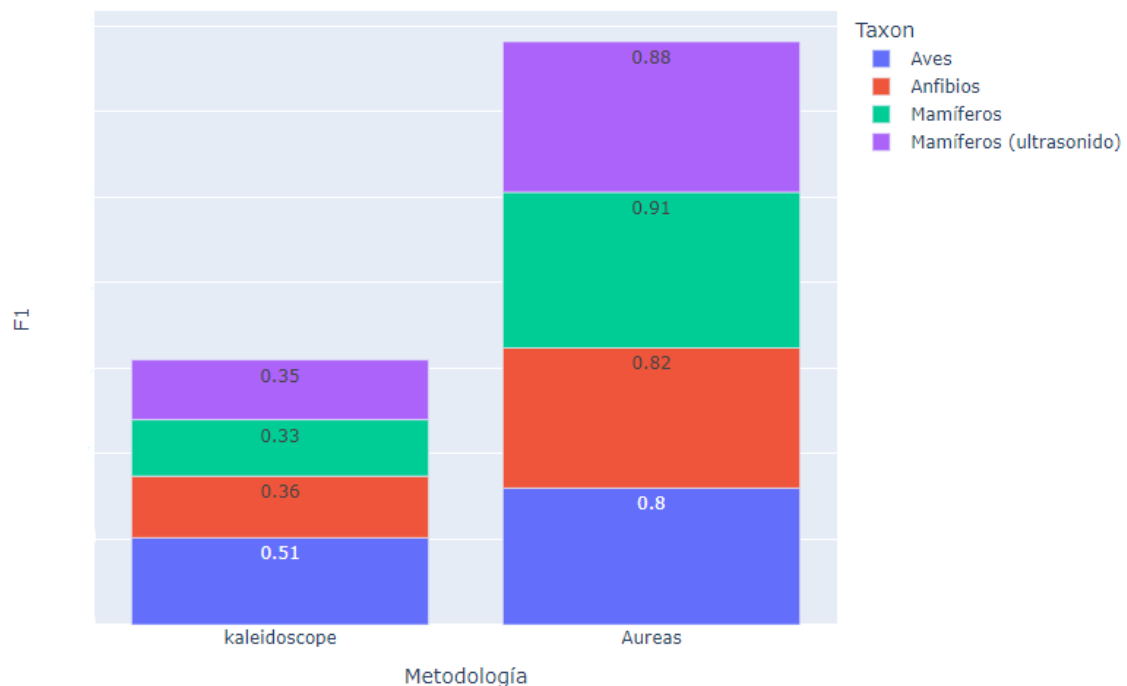


Figura 19. Compilación de resultados promedio por Taxón para cada Metodología para especies en el espectro audible

Finalmente, la figura 20 se amplía el análisis para abarcar los resultados de las especies de murciélagos en el espectro de ultrasonido. Nuevamente se observa una superioridad de los resultados obtenidos con el software Aureas en comparación con Kaleidoscope, en todos los grupos taxonómicos.



*Figura 20. Compilación de resultados promedio por Taxón para las metodologías que trabajan espectro audible y ultrasonido.*

## **Conclusiones**

Se seleccionaron y evaluaron cinco propuestas de algoritmos y softwares para la identificación automática de especies de diferentes grupos taxonómicos a través de grabaciones de audio. Durante la selección de estas metodologías, se consideraron factores como la interfaz de usuario y la facilidad de uso. Se realizaron pruebas exhaustivas con cada metodología utilizando grabaciones de audio con presencia de especies de distintos grupos taxonómicos, y se evaluó el rendimiento de cada metodología mediante la métrica F1-Score.

La evaluación de las metodologías abarcó diversidad de grupos taxonómicos presentes en las grabaciones de audio. Este enfoque reveló que la elección de una metodología debe estar en sintonía con las características particulares de las especies de interés, pues se observaron diferencias significativas en el desempeño de las metodologías en función de los grupos taxonómicos estudiados (Aves, anfibios y mamíferos). Cada grupo taxonómico presenta desafíos únicos que pueden sortearse con la elección de la herramienta adecuada.

Además, se resalta la relevancia de la optimización de interfaces de usuario, la simplificación de la configuración de parámetros y las estrategias que faciliten la reproducibilidad de los resultados, factores que pueden hacer a estas herramientas más accesibles para los expertos en biología. Un ejemplo concreto es el caso de Koogu, donde la capacidad de entrenar un clasificador a partir de las tablas de selección exportadas de Raven durante el proceso de etiquetado de la base de datos, abre la puerta a compartir la información del experimento con otros colaboradores, permitiéndoles lograr los mismos resultados y detección de la misma especie en sus propias bases de datos. De igual forma, los resultados obtenidos al utilizar el modelo de Koogu en el detector de Raven pueden utilizarse para reentrenar el modelo, logrando un modelo incluso mejor.

Los resultados de este trabajo ofrecen una guía valiosa a los investigadores para tomar decisiones informadas al seleccionar herramientas para el estudio de la biodiversidad a través del monitoreo acústico y las metodologías de reconocimiento automático de llamados. Esta guía no solo considera el rendimiento y las diferencias entre las metodologías en función de los grupos taxonómicos estudiados, sino también destaca la importancia de comprender la información de entrada requerida por cada metodología para el proceso de entrenamiento del modelo (etiquetas, audios, formatos, etc).

Conocer las especificaciones y requisitos de cada herramienta es fundamental para aprovechar al máximo las ventajas que ofrece el monitoreo acústico pasivo y las metodologías de detección automática en el estudio de la biodiversidad. Estas consideraciones, que van desde la configuración de parámetros hasta la facilidad de compartir información y mejorar los resultados, permiten a los investigadores avanzar en su trabajo de manera más efectiva y colaborativa, contribuyendo así a una mejor comprensión y conservación de la biodiversidad.

### **Referencias Bibliográficas**

- [1] Gibbs, James P.; Snell, Howard L.; Causton, Charlotte E. Effective monitoring for adaptive wildlife management: lessons from the Galapagos Islands. *The Journal of Wildlife Management*, 1999, p. 1055-1065.
- [2] Wrege, Peter H., et al. Acoustic monitoring for conservation in tropical forests: examples from forest elephants. *Methods in Ecology and Evolution*, 2017, vol. 8, no 10, p. 1292-1301.
- [3] Bergler, C., Smeele, S.Q., Tyndel, S.A. et al. ANIMAL-SPOT enables animals independent signal detection and classification using deep learning. *Sci Rep* 12, 21966 (2022). <https://doi.org/10.1038/s41598-022-26429-y>
- [4] Michaud, Félix, et al. Unsupervised classification to improve the quality of a bird song recording dataset. *Ecological Informatics*, 2023, vol. 74, p. 101952.
- [5] Ruff, Z., Lesmeister, D. B., Jenkins, J., & Sullivan, C. M. (2022). Pnw-Cnet v4: Automated species identification for passive acoustic monitoring. Available at SSRN 4167074.
- [6] Aide TM, Corrada-Bravo C, Campos-Cerqueira M, Milan C, Vega G, Alvarez R. 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1:e103 <https://doi.org/10.7717/peerj.103>
- [7] Menon, A. M., Pereira, M. J., & Aguiar, L. M. (2018). Are automated acoustic identification software reliable for bat surveys in the Neotropical region?. *PeerJ Preprints*, 6, e26712v4.
- [8] Bradfer-Lawrence, Tom, et al. Guidelines for the use of acoustic indices in environmental research. *Methods in Ecology and Evolution*, 2019, vol. 10, no 10, p. 1796-1807. <https://doi.org/10.1111/2041-210X.13254>



- [9] Xie, Jie, et al. Frog calling activity detection using lightweight CNN with multi-view spectrogram: A case study on Kroombit tinker frog. *Machine Learning with Applications*, 2022, vol. 7, p. 100202.
- [10] Bedoya, Carol L.; Molles, Laura E. Acoustic censusing and individual identification of birds in the wild. *bioRxiv*, 2021, p. 2021.10. 29.466450. Doi: <https://doi.org/10.1101/2021.10.29.466450>
- [11] Jiang, J. J., Bu, L. R., Duan, F. J., Wang, X. Q., Liu, W., Sun, Z. B., & Li, C. Y. (2019). Whistle detection and classification for whales based on convolutional neural networks. *Applied Acoustics*, 150, 169-178.
- [12] Bravo, Carlos J. Corrada; Berríos, Rafael Álvarez; Aide, T. Mitchell. Species specific audio detection: a comparison of three template-based detection algorithms using random forests. *PeerJ Computer Science*, 2017, vol. 3, p. e113.
- [13] Malfante, M., Mars, J. I., Dalla Mura, M., & Gervaise, C. (2018). Automatic fish sounds classification. *The Journal of the Acoustical Society of America*, 143(5), 2834-2846.
- [14] Astuti, W., et al. Animal sound activity detection using multi-class support vector machines. En 2011 4th international conference on mechatronics (ICOM). IEEE, 2011. p. 1-5.
- [15] Bedoya, Carol, et al. Automatic recognition of anuran species based on syllable identification. *Ecological Informatics*, 2014, vol. 24, p. 200-209.
- [16] Guerrero, M. J., Bedoya, C. L., López, J. D., Daza, J. M., & Isaza, C. (2023). Acoustic animal identification using unsupervised learning. *Methods in Ecology and Evolution*, 14, 1500–1514.
- [17] Ilyas Potamits, Stavros Ntalampiras, Olaf Jahn, Klaus Riede. (2014). Automatic bird sound detection in long real-field recordings: Applications and tools. *Applied Acoustics*, 80, 1-9.
- [18] Sugai, L. S. M., Silva, T. S. F., Ribeiro Jr, J. W., & Llusia, D. (2019). Terrestrial passive acoustic monitoring: review and perspectives. *BioScience*, 69(1), 15-25.
- [19] AviSoft Bioacoustics. (2023). SASLab Pro: User Manual version 5.2. <https://www.avisoft.com/SASLabPro.pdf>
- [20] Obrist, Martin K., et al. Bioacoustics approaches in biodiversity inventories. *Abc Taxa*, 2010, vol. 8, p. 68-99.

- [21] Menon, A. M., Pereira, M. J., & Aguiar, L. M. (2018). Are automated acoustic identification software reliable for bat surveys in the Neotropical region?. *PeerJ Preprints*, 6, e26712v4.
- [22] Ren, Yao, et al. A framework for bioacoustic vocalization analysis using hidden Markov models. *Algorithms*, 2009, vol. 2, no 4, p. 1410-1428. Doi: <https://doi.org/10.3390/a2041410>
- [23] LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J. P., Dodhia, R., Ferres, J. L., & Aide, T. M. (2020). A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. *Ecological Informatics*, 59, 101113.
- [24] Aide, T. M., Hernández-Serna, A., Campos-Cerqueira, M., Acevedo-Charry, O., & Deichmann, J. L. (2017). Species richness (of insects) drives the use of acoustic space in the tropics. *Remote Sensing*, 9(11), 1096.
- [25] Monroe, K. D., Collazo, J. A., Pacifici, K., Reich, B. J., Puente-Rolón, A. R., & Terando, A. J. (2017). Occupancy and abundance of *Eleutherodactylus* frogs in coffee plantations in Puerto Rico. *Herpetologica*, 73(4), 297-306.
- [26] Zhong, M., LeBien, J., Campos-Cerqueira, M., Dodhia, R., Ferres, J. L., Velev, J. P., & Aide, T. M. (2020). Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling. *Applied Acoustics*, 166, 107375.
- [27] Guerrero, M.J., Restrepo, J., Nieto-Mora, D.A., Daza, J.M., Isaza, C. (2022). Insights from Deep Learning in Feature Extraction for Non-supervised Multi-species Identification in Soundscapes. In: Bicharra Garcia, A.C., Ferro, M., Rodríguez Ribón, J.C. (eds) *Advances in Artificial Intelligence – IBERAMIA 2022. Lecture Notes in Computer Science()*, vol 13788. Springer, Cham. [https://doi.org/10.1007/978-3-031-22419-5\\_19](https://doi.org/10.1007/978-3-031-22419-5_19)
- [28] Restrepo Rivas, J. (2022). Redes neuronales convolucionales como alternativa en la extracción de características para la identificación automática de especies a partir de audios. Trabajo de Investigación, Universidad de Antioquia, Medellín, Colombia.
- [29] Wildlife Acoustics, "Kaleidoscope Pro 5 User Guide", Wildlife Acoustics, Inc., 2018. <http://condor.wildlifeacoustics.com/Kaleidoscope.pdf>
- [30] Miller, B. S., Madhusudhana, S., Aulich, M. G., & Kelly, N. (2023). Deep learning algorithm outperforms experienced human observer at detection of blue whale D-calls:

a double-observer analysis. *Remote Sensing in Ecology and Conservation*, 9(1), 104-116.

## Anexos

### *Anexo 1: Recopilación de bibliografía.*

Nombre	Uso	Grupos taxonómicos trabajados	GUI
Low-cost open-source recorders and ready-to-use machine learning approaches provide effective monitoring of threatened species	Se utilizó kaleidoscope pro para generar etiquetas de presencia/ausencia y compararlas con unas manualmente anotadas	Aves	Si
Software performance for the automated identification of bird vocalisations: the case of two closely related species	Detección, segmentación y clasificación de eventos acústicos	Aves	Si
Are automated acoustic identification software reliable for bat surveys in the Neotropical region?	Identificación automática	Mamíferos en ultrasonido (murciélagos)	Si
Using Acoustic Monitoring to Detect Bat Species in the University of Idaho Experimental Forest	Identificación automática	Mamíferos en ultrasonido (murciélagos)	Si
The Problem of Low Agreement among Automated Identification Programs for Acoustical Surveys of Bats	Identificación automática	Mamíferos en ultrasonido (murciélagos)	Si
Evaluation of Acoustic Recording Devices to Survey Northern Bobwhite Populations	Clasificador	Aves	Si
Species detection framework using automated recording units: a case study of the Critically Endangered Jerdon's courser	Identificación automática	Aves	Si
Workflow and convolutional neural network for automated identification of animal sounds	Etiquetado con la herramienta de clustering, y construcción de una herramienta de clustering con CNN	Aves y mamíferos	Si
Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling	Construcción de la base de datos con la herramienta template matching	Aves, anfibios	No
Real-time bioacoustics monitoring and automated species identification	Creación de modelos especie-específicos para anfibios, aves, mamíferos e insectos	Anfibios, aves, mamíferos e insectos	Si
A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network	Utilizaron arbimon para el almacenamiento, visualización y creación de una base de datos de entrenamiento utilizando la herramienta de pattern matching.	Aves, anfibios	No
Climate change is creating a	Se utilizó arbimon para almacenar,	Aves, anfibios	Si

mismatch between protected areas and suitable habitats for frogs and birds in Puerto Rico	hacer pattern matching y analizar presencia y ausencia, verificada luego manualmente		
Species Richness (of Insects) Drives the Use of Acoustic Space in the Tropics	Se almacenaron los audios en la plataforma y se utilizó la herramienta de análisis de paisaje sonoro	Anfibios, peces, aves e insectos	Si
Occupancy and Abundance of Eleutherodactylus Frogs in Coffee Plantations in Puerto Rico	Almacenamiento y análisis de los datos con arbimon y random forest	Anfibios	Si
Applications and advances in acoustic monitoring for infectious disease epidemiology	Análisis de frecuencia de detección de especies en los sitios de investigación	Anfibios, insectos, aves y primates	Si
Passive Acoustic Monitoring as a Tool to Investigate the Spatial Distribution of Invasive Alien Species	Se almacenaron los datos para hacer un modelo de ocupación de especies invasoras	Anfibios, aves y mamíferos	Si
Software performance for the automated identification of bird vocalisations: the case of two closely related species	Band limited energy detector para la detección de eventos de audio.	Aves	Si
Bioacoustics: The International Journal of Animal Sound and its Recording	Se utilizó la herramienta de correlación de raven pro para encontrar la diferencia de tiempo de aparición de las vocalizaciones en un arreglo de micrófonos	Aves, anfibios	Si
Species detection framework using automated recording units: a case study of the Critically Endangered Jerdon's courser	Identificación automática	Aves	Si
Are automated acoustic identification software reliable for bat surveys in the Neotropical region?	Se hicieron identificaciones automáticas para luego hacer verificación manual	Mamíferos en ultrasonido (murciélagos)	Si
Are automated acoustic identification software reliable for bat surveys in the Neotropical region?	Identificación de presencia	Mamíferos en ultrasonido (murciélagos)	Si
TRITON software package: Analyzing large passive acoustic monitoring data sets using MATLAB	Análisis de grandes bases de datos de monitoreo acústico pasivo	-	Si
DetEdit: A graphical user interface for annotating and editing events detected in long-term acoustic monitoring data	Análisis de grandes bases de datos de monitoreo acústico pasivo	Mamíferos	Si
PNW-Cnet v4: Automated Species Identification for Passive Acoustic Monitoring	Se creó un sistema para la detección de múltiples especies	Aves, anfibios, mamíferos e insectos	Si
Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring	Estimación de población con detección automática	Mamíferos	Si
Automatic Detection and Compression for Passive Acoustic Monitoring of the African Forest Elephant	Se creó un sistema para el monitoreo de Elefante africano	Mamíferos	No
A review of current marine mammal	Revisión de algoritmos para	Mamíferos	No

detection and classification algorithms for use in automated passive acoustic monitoring.	detección de mamíferos marinos	marinos	
Passive Acoustic Monitoring and Automatic Detection of Diel Patterns and Acoustic Structure of Howler Monkey Roars	Se creó un sistema para el monitoreo de Monos	Mamíferos	No
Passive acoustic monitoring as a potential tool to survey animal and ecosystem processes in freshwater environments	Análisis del estado del ecosistema a partir de la detección de especies	Anfibios	No
Three-dimensional passive acoustic tracking of sperm whales ( <i>Physeter macrocephalus</i> ) in ray-refracting environments	Se creó un sistema para el monitoreo de cachalotes	Mamíferos marinos	Si
Open-source workflow approaches to passive acoustic monitoring of bats	Clasificación de especies de murciélagos, comparando diversos métodos de detección y clasificación	Mamíferos en ultrasonido (murciélagos)	Si
A Permanent Automated Real-Time Passive Acoustic Monitoring System for Bottlenose Dolphin Conservation in the Mediterranean Sea	Sistema de detección automática de mamíferos marinos	Mamíferos marinos	Si
An Industry Perspective of Using Towed Passive Acoustic Monitoring (PAM) for the Detection of Marine Mammals at Sea During Seismic Surveys	Sistema de detección automática de mamíferos marinos	Mamíferos marinos	Si
Tadarida: A Toolbox for Animal Detection on Acoustic Recordings	Construcción de un software para detección y clasificación automática	Mamíferos en ultrasonido (murciélagos) e insectos	Si
Common humpback whale ( <i>Megaptera novaeangliae</i> ) sound types for passive acoustic monitoring	Monitoreo de ballenas jorobadas	Mamíferos	Si
Automatic fish sounds classification	Clasificación de sonidos de peces	Peces	No
soundClass: An automatic sound classification tool for biodiversity monitoring using machine learning	Software de detección y clasificación automática	Mamíferos en ultrasonido (murciélagos)	Si
Whistle detection and classification for whales based on convolutional neural networks	Sistema de detección y clasificación para 2 especies de ballenas	Mamíferos marinos	Si
Automated classification of dolphin echolocation click types from the Gulf of Mexico	Detección y clasificación de clicks de delfines	Mamíferos marinos	Si