Reducción de ruido en señales bioacústicas: Un enfoque basado en wavelets y aplicado al monitoreo de poblaciones de aves y anfibios

Estudiante:

Sebastian Carvalho Salazar, ✉ scarvalhos@eafit.edu.co

Asesor:

Johan Felipe García Vargas, ✉ jfgarcaav@eafit.edu.co

Universidad EAFIT
Maestría en Ciencia de Datos y Analítica
Medellín
2025

Resumen

Las técnicas de monitoreo bioacústico permiten la detección no invasiva de poblaciones biológicas mediante grabadoras automáticas que capturan de manera continua vocalizaciones de especies en su entorno natural. Este estudio evalúa el impacto de técnicas de reducción de ruido basadas en transformadas wavelet sobre el procesamiento de señales bioacústicas y su influencia en modelos de clasificación de referencia, como BirdNET para aves y AnuraSet para anfibios. La metodología incluye una fase de preprocesamiento de reducción de ruido y un análisis detallado de métricas de rendimiento, como la similitud coseno mel, la correlación temporal, el indice de entropía, la confianza y las curvas ROC-AUC. Los resultados muestran que la reducción de ruido mejora la claridad de la señal y reduce tasas de falsas alarmas, facilitando una discriminación más precisa en entornos acústicamente complejos como áreas urbanas y selvas. Aunque la técnica suprime ciertas características sutiles de las vocalizaciones, el análisis estadístico y los gráficos radar sugieren que con ajustes en el proceso de denoising es posible optimizar el balance entre reducción de ruido y preservación de características bioacústicas esenciales. Por tanto, la reducción de ruido mediante wavelets es una estrategia viable y robusta en entornos de alta interferencia, aunque podría no ser óptima para estudios que requieren la captura exhaustiva de todas las vocalizaciones, como en especies en peligro o de baja densidad. Además, el denoising regula la confianza en predicciones incorrectas y preserva las características relevantes en predicciones correctas, reduciendo falsas alarmas.

Palabras clave: Monitoreo bioacústico, Reducción de ruido, Wavelets, Señales bioacústicas, BirdNET, AnuraSet.

Abstract

Bioacoustic monitoring techniques enable non-invasive detection of biological populations through automatic recorders that continuously capture species vocalizations in natural habitats. This study assesses the impact of wavelet-based noise reduction on bioacoustic signal processing and evaluates its influence on benchmark classification models, specifically BirdNET for birds and AnuraSet for amphibians. Our methodology includes noise reduction preprocessing, followed by an in-depth analysis of classification performance metrics such as mel cosine similarity, temporal correlation, entropy ratio, and ROC-AUC curves. Results indicate that noise reduction enhances signal clarity and reduces false alarm rates, enabling more accurate discrimination in acoustically complex environments like urban areas and rainforests. Although the technique may suppress some subtle vocalization features, statistical analysis and radar plots suggest that adjustments to the denoising process can help optimize the balance between noise reduction and preservation of essential bioacoustic characteristics. Consequently, wavelet-based noise reduction is a robust strategy for high-interference environments, though it may be less suitable for studies requiring comprehensive capture of all vocalizations, such as endangered or low-density species. Moreover, denoising regulates confidence in incorrect predictions and preserves relevant features in correct predictions, reducing false alarms.

Keywords: Bioacoustic monitoring, Noise reduction, Wavelets, BirdNET, AnuraSet.

TABLA DE CONTENIDOS

I. Introducción 1

II. Justificación 3

III. Objetivos 5
- III-A. Objetivo general . . . 5
- III-B. Objetivos específicos . . . 5

IV. Marco Teórico 6
- IV-A. Procesamiento de señales . . . 6
  - IV-A1. Tasa de Bits . . . 7
  - IV-A2. Teorema de Nyquist . . . 7
  - IV-A3. Filtro Butterworth . . . 7
  - IV-A4. Descomposición Wavelet . . . 8
  - IV-A5. Umbralización de Coeficientes Wavelet . . . 9
  - IV-A6. SureShrink . . . 10
  - IV-A7. BayesShrink . . . 10
  - IV-A8. Selección del Nivel Óptimo de Wavelet . . . 12
  - IV-A9. Correlación de Pearson . . . 13
  - IV-A10. Overfitting en wavelets . . . 14
  - IV-A11. Similitud Coseno de los Coeficientes Mel . . . 14
  - IV-A12. Ratio de Entropía . . . 15
- IV-B. Modelos de aprendizaje estadístico . . . 16
  - IV-B1. Modelos de Clasificación de Especies . . . 16
  - IV-B2. Metodología de evaluación . . . 17
  - IV-B3. Gráficos Radar y Pruebas de Hipótesis . . . 18
  - IV-B4. Curvas ROC y Evaluación del AUC . . . 18
  - IV-B5. Pruebas de Hipótesis (AUC) . . . 19
  - IV-B6. Distancia de Jaccard . . . 20
  - IV-B7. Métrica utilizada para evaluar la efectividad de cada wavelet 20

V. Metodología 21
- V-A. Comprensión del negocio . . . 21
- V-B. Comprensión de los datos . . . 21
  - V-B1. Parex . . . 22
  - V-B2. AnuraSet . . . 22
- V-C. inspección visual y auditiva de los datos . . . 23
- V-D. Preparación de los datos . . . 24
- V-E. Modelado . . . 24
- V-F. Evaluación . . . 25
- V-G. Despliegue . . . 25

VI. Resultados y discusión 26
- VI-A. Ruido vs. Vocalizaciones de Especies . . . 28
  - VI-A1. Gráficos Radar: Evaluación de Métricas de Denoising . . 28
  - VI-A2. Gráfico de Barras de Proporciones: . . . 30

VI-A3. Análisis de la Distancia de Jaccard 30
VI-A4. Análisis de la Curva ROC 31
VI-B. Vocalizaciones de Especies Frecuentes vs. Ruido 32
VI-B1. Resultados para *Atalotriccus pilaris* 33
VI-B2. Gráfico de Barras de Proporciones 34
VI-B3. Resultados para *Pitangus sulphuratus* 35
VI-B4. Gráfico de Barras de Proporciones 36
VI-B5. Resultados para *Tolmomyias flaviventris* 37
VI-B6. Gráfico de Barras de Proporciones 38
VI-B7. Análisis de la Curva ROC 38
VI-C. Ruido vs. Vocalizaciones de Especies 41
VI-C1. Gráficos Radar: Evaluación de Métricas de Denoising . . 42
VI-C2. Gráfico de Barras de Proporciones 43
VI-C3. Análisis de la Distancia de Jaccard 44
VI-C4. Análisis de la Curva ROC y Distribuciones de AUC . . . 44
VI-D. Vocalizaciones de Especies Frecuentes vs. Ruido 46
VI-D1. Resultados para *DENMIN* 46
VI-D2. Gráfico de Barras de Proporciones 48
VI-D3. Resultados para *LEPPOD* 48
VI-D4. Gráfico de Barras de Proporciones 50
VI-D5. Resultados para *PHYALB* 50
VI-D6. Gráfico de Barras de Proporciones 52
VI-D7. Análisis de la Curva ROC 52

VII. Conclusiones 55

VIII. Plan de Gestión de Datos 57

IX. Aspectos Éticos 58

X. Trabajos Futuros 59

XI. Anexos 61
XI-A. Interfaz de Usuario y Funcionamiento de la Aplicación 61
XI-B. Procedimiento para Solicitud de Acceso bajo Confidencialidad 62

XII. Referencias 63

LISTA DE FIGURAS

1. Visualización de la respuesta en frecuencia de un filtro pasa banda Butterworth de orden 5, con un rango de frecuencias de 500 Hz a 4000 Hz. 8
2. Columbina squammata. Formas de onda de las señales de audio originales y denoised aplicando diferentes métodos de umbral. Se utilizó el wavelet **dmey** con nivel 9 en todos los casos: **VisuShrink**: Correlación: 0.7, Similitud coseno (Mel): 0.932, Ratio de entropía: 0.898. **SureShrink**: Correlación: 0.764, Similitud coseno (Mel): 0.944, Ratio de entropía: 0.918. **BayesShrink**: Correlación: 0.776, Similitud coseno (Mel): 0.944, Ratio de entropía: 0.928. 11
3. Columbina squammata. Espectrogramas de las señales de audio originales y denoised aplicando diferentes métodos de umbral. Se utilizó el wavelet **dmey** con nivel 9 en todos los casos: **VisuShrink**: Correlación: 0.7, Similitud coseno (Mel): 0.932, Ratio de entropía: 0.898. **SureShrink**: Correlación: 0.764, Similitud coseno (Mel): 0.944, Ratio de entropía: 0.918. **BayesShrink**: Correlación: 0.776, Similitud coseno (Mel): 0.944, Ratio de entropía: 0.928. 11
4. Visualización comparativa de los métodos de selección del nivel óptimo de descomposición wavelet. Se muestran dos enfoques: el criterio de aumento de entropía y el método del codo. Los datos representados son sintéticos, generados para ejemplificar los métodos. 13
5. Comparación de las señales bioacústicas antes y después de la reducción de ruido utilizando el método VisuShrink. Se presentan la forma de onda en el dominio del tiempo, la Densidad Espectral de Potencia (PSD) y los espectrogramas Mel para diferentes especies, permitiendo analizar el comportamiento del filtrado y la preservación de características acústicas clave. 23
6. Top 12 wavelets en los escenarios de ”Noise + Specie Event Audios”(izquierda) y ”Noise Audios”(derecha). El valor de la métrica se calcula como el producto entre la mediana de la confianza asignada por el modelo BirdNET y el número de veces que cada wavelet fue seleccionada como la mejor en el proceso de eliminación de ruido. Las barras incluyen errores calculados como MAD (Median Absolute Deviation) para la confianza. 26
7. Frecuencia y proporción de selección de cada método de umbralización (bayesshrink, sureshrink, visushrink) en el proceso de eliminación de ruido. . . . 27
8. Gráficos radar de métricas de evaluación para casos con predicción correcta (derecha) e incorrecta (izquierda). Los gráficos incluyen los intervalos de confianza del 95 % para cada métrica y una referencia de confianza en el caso sin denoising. 29
9. Proporción de predicciones correctas y erróneas para los tipos de audio “Audio Original” y “Audio Denoised” (679 muestras). 30
10. Curvas ROC con intervalos de confianza para los casos Dirty (Audio Original) y Denoised (Eliminación de Ruido). La curva en negro representa el rendimiento del modelo con señales originales, con un AUC de 0.687, y su intervalo de confianza al 95 % sombreado en gris. La curva en rojo muestra el rendimiento del modelo con señales denoised, alcanzando un AUC de 0.726, con su intervalo de confianza al 95 % sombreado en rojo. La mejora en el AUC sugiere que el denoising facilita la clasificación correcta al reducir la interferencia del ruido ambiental. 31

11. Histogramas de AUC obtenidos mediante bootstrapping para los modelos denoised y dirty (Audios sin procesar). Las líneas punteadas representan las medianas para cada caso: 0.727 para denoised y 0.687 para dirty. 32
12. *Atalotriccus pilaris* - Comparación de métricas para predicciones correctas e incorrectas, incluyendo intervalos de confianza del 95 %. 33
13. Proporción de clasificaciones correctas e incorrectas para *Atalotriccus pilaris* en los audios originales y procesados con denoising (402 Muestras). 34
14. *Pitangus sulphuratus* - Comparación de métricas para predicciones correctas e incorrectas, incluyendo intervalos de confianza del 95 %. 35
15. Proporción de clasificaciones correctas e incorrectas para *Pitangus sulphuratus* en los audios originales y procesados con denoising (440 Muestras). 36
16. *Tolmomyias flaviventris* - Comparación de métricas para predicciones correctas e incorrectas, incluyendo intervalos de confianza del 95 %. 37
17. Proporción de clasificaciones correctas e incorrectas para *Tolmomyias flaviventris* en los audios originales y procesados con denoising (384 Muestras). . 38
18. Comparación de curvas ROC y distribuciones AUC para diferentes especies: Atalotriccus pilaris, Pitangus sulphuratus y Tolmomyias flaviventris. Cada pareja muestra la curva ROC junto con el histograma de AUC para cada especie, comparando datos originales y denoised. 39
19. Top 12 wavelets seleccionadas en los experimentos con anuros para diferentes condiciones de ruido. 40
20. Frecuencia y proporción de selección de cada método de umbralización (bayesshrink, sureshrink, visushrink) en el proceso de eliminación de ruido en señales de anuros. 41
21. Gráfico radar de las métricas de evaluación para predicciones correctas e incorrectas en el caso de los anuros, mostrando la similitud coseno en el dominio Mel, la correlación temporal, la razón de entropía y la confianza del modelo. 42
22. Proporción de predicciones correctas e incorrectas por tipo de audio (Audio Denoised y Audio Original) en el caso de los anuros (523 Muestras). 43
23. Curva ROC con intervalos de confianza para los casos Dirty (Señal Origianal) y Denoised (Señal Procesada). El AUC mediano es mayor para el audio original (0.919) en comparación con el audio denoised (0.852), lo que sugiere un impacto negativo del proceso de denoising en la precisión de clasificación. . . 45
24. Distribuciones de los valores de AUC obtenidos mediante bootstrap para los casos con y sin eliminación de ruido. Se observa que el AUC mediano es mayor para el audio original (0.919) en comparación con el audio denoised (0.850), lo que indica una mayor precisión de clasificación sin el proceso de reducción de ruido. 46
25. Gráficos de radar de las métricas de evaluación para las predicciones incorrectas (izquierda) y correctas (derecha) en la especie *DENMIN*. 47
26. Proporción de predicciones correctas y erróneas para la especie *DENMIN* en audios denoised y originales (640 Muestras). 48
27. Gráficos de radar de las métricas de evaluación para las predicciones incorrectas (izquierda) y correctas (derecha) en la especie *LEPPOD*. 49
28. Proporción de predicciones correctas y erróneas para la especie *LEPPOD* en audios denoised y originales (842 Muestras). 50

29. Gráficos de radar de las métricas de evaluación para las predicciones incorrectas (izquierda) y correctas (derecha) en la especie *PHYALB*. 51
30. Proporción de predicciones correctas y erróneas para la especie *PHYALB* en audios denoised y originales (434 Muestras). 52
31. Comparación de curvas ROC y distribuciones AUC para diferentes especies de anuros: DENMIN, LEPPOD y PHYALB. Cada pareja muestra la curva ROC junto con el histograma de AUC para cada especie, comparando datos originales y denoised. 54
32. Interfaz gráfica de la aplicación para denoising de audio con Wavelets en Gradio 61

I. Introducción

El monitoreo bioacústico es una técnica de creciente interés en la ecología y biología de la conservación, ya que permite registrar la actividad vocal de diversas especies de manera no invasiva. Esta metodología se basa en la captura de señales acústicas del entorno, proporcionando una herramienta efectiva para la evaluación de la biodiversidad y la detección de especies en hábitats naturales. Su uso ha demostrado ser crucial para estudios de ecología del paisaje, monitoreo de poblaciones y evaluación del impacto ambiental [1], [2]. El principal desafío en la captura pasiva de señales bioacústicas radica en la interferencia del ruido ambiental, lo que afecta significativamente la relación señal-ruido y, en consecuencia, el desempeño de los modelos de aprendizaje profundo en comparación con grabaciones obtenidas mediante micrófonos direccionales operados por humanos. Esta contaminación acústica introduce componentes espectro-temporales irrelevantes, dificultando la extracción de características discriminantes esenciales para la clasificación de especies.

Este proyecto aborda la evaluación del impacto de las técnicas de reducción de ruido basadas en wavelets en el procesamiento de señales bioacústicas y su influencia en el rendimiento de modelos de referencia para la clasificación de datos de este tipo. Las señales, obtenidas mediante dispositivos especializados, sufren contaminación por ruido ambiental y artefactos no deseados, lo que dificulta su análisis y la extracción de características relevantes para la identificación de especies. Este desafío se ve abordado por el uso de modelos de aprendizaje automático como: BirdNET y AnuraSet, herramientas cruciales en la supervisión de la riqueza biológica. El uso de datos acústicos recolectados de manera pasiva para el monitoreo de la diversidad ofrece varios beneficios potenciales, como el monitoreo costo-efectivo, alta resolución temporal y reducción de sesgos [1]. Estas ventajas son fundamentales para mejorar los esfuerzos de conservación e investigación ecológica al proporcionar un enfoque para ello.

BirdNET, desarrollado mediante redes neuronales profundas con arquitectura ResNet, puede identificar 984 especies de aves de América del Norte y Europa a partir de grabaciones de sonido, lo que resalta su utilidad en contextos donde la diversidad aviar es alta [1]. Por otro lado, AnuraSet, que está diseñado específicamente para la identificación de anuros, emplea una red neuronal basada en la arquitectura ResNet, entrenada con un conjunto de datos anotado de vocalizaciones de anuros neotropicales, facilitando la identificación automática de estas especies [3], [4]. Ambas herramientas han demostrado su eficacia en la clasificación automática de especies a partir de grabaciones de sonido, siendo indicadores efectivos para el monitoreo eficiente de la vida silvestre en una variedad de entornos.

Las aves y los anfibios presentan notables diferencias acústicas y desempeñan roles distintos como indicadores ambientales. Las aves se caracterizan por una amplia variedad de vocalizaciones, con frecuencias y patrones complejos que varían significativamente entre

especies. Por otro lado, los anfibios, como las ranas, producen vocalizaciones más simples y especializadas, aunque algunas especies pueden utilizar un rango de frecuencias más amplio. En cuanto a la superposición de poblaciones, los anfibios suelen concentrarse en hábitats acuáticos durante la época de reproducción, lo que genera una mayor densidad en estas áreas. Por su parte, ciertas especies de aves pueden mostrar una alta superposición temporal en momentos específicos, como durante la migración o en zonas con abundantes recursos [3]. Aunque las aves han sido estudiadas con mayor profundidad, los anfibios, debido a su sensibilidad a los cambios ambientales, representan indicadores excepcionales de la salud de los ecosistemas. En el contexto colombiano, un enfoque integral que contemple el monitoreo de ambos grupos como indicadores ambientales ofrecería una perspectiva más completa para comprender y conservar la biodiversidad y la salud de los ecosistemas en este país megadiverso.

Los resultados obtenidos en este estudio demuestran el impacto de las técnicas de reducción de ruido mediante wavelets en la precisión y confianza de modelos de clasificación bioacústica, específicamente BirdNET y AnuraSet. Las pruebas realizadas indican que, si bien el denoising basado en wavelets logra reducir significativamente el ruido ambiental, en algunos casos puede comprometer la precisión de la clasificación al eliminar detalles acústicos relevantes. No obstante, el uso de wavelets seleccionadas y métodos de umbralización como *sureshrink* demostró una efectividad notable en la conservación de las características bioacústicas clave en entornos complejos. Los resultados y comparaciones obtenidos en este trabajo resaltan la importancia de un equilibrio adecuado entre la reducción de ruido y la preservación de señales bioacústicas.

II. Justificación

La investigación en el análisis de señales acústicas es esencial para el monitoreo de poblaciones biológicas, permitiendo la clasificación automatizada de especies sin necesidad de contacto directo. Aunque la utilización de micrófonos direccionales es común en la captura de fuentes sonoras en un ecosistema, la evolución hacia dispositivos que permiten la adquisición de datos con una menor intervención humana, como los sistemas de internet de las cosas (IoT), ofrecen ventajas al capturar un volumen más amplio [5], [6]. Sin embargo, hay una diferencia fundamental en la captura de estos datos que juega un papel crucial en el proceso de eliminación de ruido.

Cuando se utilizan micrófonos direccionales, la señal de interés se concentra y la relación señal a ruido suele ser mejor, lo que puede hacer que la aplicación de técnicas de reducción de ruido tenga el riesgo de eliminar información importante de la especie en particular, al ser ésta la componente dominante del audio. Por otro lado, en la captura con micrófonos no dirigidos u omnidireccionales, la especie de interés se mezcla con su entorno natural y se captura todo el ruido ambiental circundante además de la señal de interés [7]. En este escenario, la aplicación de técnicas wavelet para reducción de ruido puede ser beneficiosa, ya que su capacidad para procesar información tiempo-frecuencia puede permitir la discriminación entre la señal de interés y el ruido ambiental, preservando así la información esencial sobre la especie en cuestión.

El enfoque tradicional de eliminación de ruido se basa en la aplicación de filtros, que son efectivos cuando la señal y el ruido están en bandas de frecuencia diferentes. Sin embargo, este método es incapaz de preservar la señal deseada cuando el contenido espectral se superpone. La transformada Wavelet ha sido ampliamente utilizada en el campo del procesamiento de señales digitales debido a su capacidad para procesar características tiempo-frecuencia, minimizando la pérdida de información significativa, lo que supera al análisis de Fourier tradicional permitiendo detectar cambios locales en señales no estacionarias, lo que mejora la preservación de información relevante [8].

A pesar de estos avances, es importante destacar que los modelos de redes neuronales como BirdNET y AnuraSet no han incorporado la reducción de ruido en su pipeline de procesamiento. Nuestro objetivo es entender cómo estas metodologías impactan el desempeño de modelos entrenados con técnicas de data augmentation avanzadas, como BirdNET, cuyo entrenamiento ha sido diseñado para robustecer el modelo ante desplazamientos en tiempo, frecuencia, enmascaramientos y ruidos ambientales. Se busca evaluar cómo estas técnicas de reducción de ruido, como las basadas en wavelets, afectan el rendimiento de modelos altamente optimizados y cómo los modelos que han sido entrenados con conjuntos de datos menos extensivos o menos robustos frente a perturbaciones acústicas podrían mostrar variaciones en su rendimiento. En estos casos, existe el riesgo de que los modelos menos

robustos, entrenados en entornos específicos, hayan desarrollado un sobreaprendizaje del ruido particular de esos entornos, lo que podría hacer que pequeñas perturbaciones inducidas por la eliminación de ruido afecten su desempeño. Esta evaluación es esencial para mejorar la calidad de los registros sonoros biológicos empleados en la vigilancia de poblaciones de aves y anfibios, integrando técnicas de procesamiento digital de señales y aprendizaje automático en un pipeline de procesamiento que puede tener consecuencias significativas para la preservación de la diversidad biológica.

En el caso específico de Colombia, un país megadiverso, la aplicación de estas técnicas puede beneficiar significativamente el monitoreo y conservación de especies tanto de aves como de anfibios [9]. Los anfibios, en particular, complementan a las aves como indicadores de calidad ambiental debido a su alta sensibilidad a los cambios en el entorno, especialmente en los ecosistemas acuáticos. Sin embargo, su estudio presenta desafíos adicionales debido a la menor atención histórica que han recibido en comparación con las aves. Además, la clasificación de datos bioacústicos en anfibios representa un reto considerable, ya que sus vocalizaciones abarcan una diversidad de señales acústicas que, en algunos casos, pueden ser complejas y difíciles de distinguir. Este desafío se ve agravado por el hecho de que el rango de frecuencias de las vocalizaciones de muchos anfibios a menudo se solapa con el ruido ambiental de sus hábitats, como el sonido del agua, el viento o las vocalizaciones de otros animales, lo que enmascara las señales acústicas y complica aún más su detección y análisis.

III. Objetivos

III-A. Objetivo general

Evaluar el impacto de las técnicas de reducción de ruido basadas en wavelets en el procesamiento de señales bioacústicas, determinando su efecto en el rendimiento de los modelos BirdNET y AnuraSet para la clasificación de especies. Para ello, se utilizará una muestra representativa de audios omnidireccionales de aves y anfibios registrados en Parex 2023 y AnuraSet (INCT17), considerando su relevancia en el estudio de la biodiversidad colombiana.

III-B. Objetivos específicos

- Desarrollar un algoritmo de reducción de ruido basado en wavelets para señales bioacústicas que se ajuste a las características específicas de la muestra de aves o anfibios.

- Analizar el impacto de la reducción de ruido en las métricas de los modelos utilizados para clasificación de datos bioacústicos.

IV. Marco Teórico

La clasificación de especies a partir de grabaciones de audio es un campo de estudio en constante crecimiento, ampliamente beneficiado por el uso de herramientas como el procesamiento digital de señales y el aprendizaje automático [1], [2]. Estas técnicas han demostrado ser útiles para abordar problemas de clasificación de especies de aves y anfibios a partir de datos acústicos, ofreciendo una manera eficaz de analizar y comprender la biodiversidad presente en diferentes entornos sonoros [10].

Las señales bioacústicas analizadas en este estudio son capturadas mediante sensores especializados, como micrófonos, que convierten la onda de presión del aire en una señal eléctrica a través de un proceso de transducción electroacústica. Inicialmente, en el dominio analógico (continuo), la perturbación de una membrana por la presión sonora genera un movimiento mecánico, el cual induce una corriente en una bobina suspendida en un campo magnético, transformando la energía mecánica en una señal eléctrica continua. Finalmente, esta señal es muestreada y digitalizada, permitiendo su procesamiento en el dominio digital (discreto) mediante técnicas computacionales avanzadas.

En este estudio, se exploran técnicas clave utilizadas en este ámbito, con un enfoque particular en incorporar la reducción de ruido a la clasificación de especies. Es decir, integrar técnicas avanzadas de reducción de ruido basadas en descomposición wavelet con modelos de clasificación profunda, con el objetivo de investigar cómo estas técnicas afectan el rendimiento de los modelos utilizados.

Por otro lado, se destaca la importancia de disponer de conjuntos de datos anotados que sirvan como punto de referencia para la identificación automatizada de especies en el monitoreo acústico pasivo. Como ejemplo, Cañas et al. (2023) presentan un conjunto de datos para el benchmarking de la identificación de llamadas de anuros neotropicales en el monitoreo acústico pasivo [3]. Este conjunto de datos proporciona una valiosa herramienta para la evaluación de modelos de identificación automatizada en entornos acústicos neotropicales, contribuyendo así al avance de la investigación en este campo.

A continuación, se explorarán diversos componentes teóricos fundamentales en el procesamiento de señales y el aprendizaje automático, con el fin de profundizar en la comprensión y aplicación de estos, tomando como base el módulo de eliminación de ruido `AudioDenoiser` implementado en este proyecto para el tratamiento de señales bioacústicas.

IV-A. Procesamiento de señales

El procesamiento de señales, en el contexto de este trabajo, emerge como un componente fundamental en el análisis y la manipulación de datos provenientes de diversas fuentes,

en este caso acústicas. Este campo de estudio, esencial en la ingeniería y la ciencia, se encarga de aplicar algoritmos computacionales para analizar, filtrar, modificar y extraer información valiosa de las señales capturadas. Desde la detección de patrones en señales biológicas hasta la optimización de sistemas de comunicación, el procesamiento de señales desempeña un papel crítico en numerosas disciplinas. Para una mayor comprensión del tema consulte [11], [12].

IV-A1) Tasa de Bits: Para convertir una señal de audio de formato `INT16` a `FLOAT32` y garantizar una correcta normalización, es esencial comprender cómo se mapean los datos en ambas representaciones. En `INT16`, los valores de las muestras de audio oscilan entre -32768 y 32767, representando cada valor en un espacio de 16 bits donde el bit más significativo indica el signo y los restantes conforman la mantisa, proporcionando la precisión de la muestra. Para ajustar esta señal al formato `FLOAT32`, que requiere una escala en el rango de $[-1, 1]$, cada muestra debe ser dividida por 2^{15} o 32768, como se expresa en la ecuación:

$$\text{signal}_{\text{float}} = \frac{\text{signal}_{\text{int16}}}{2^{15}} \tag{1}$$

Este proceso asegura que la mantisa en `FLOAT32` capture los valores precisos de la señal original, mientras que el bit de signo mantiene la polaridad de cada muestra.

IV-A2) Teorema de Nyquist: El teorema de Nyquist, un pilar en la teoría de la comunicación, establece la relación crucial entre la tasa de muestreo y la frecuencia de la señal para una reconstrucción precisa, este principio, ampliamente aplicado en campos como la telegrafía y la digitalización de señales, continúa siendo fundamental desde su formulación en 1928 [13].

$$f_s \geq 2 \cdot f_{\text{max}} \tag{2}$$

IV-A3) Filtro Butterworth: En el contexto del problema de clasificación de especies a partir de grabaciones de audio, se puede aplicar un filtro Butterworth pasa banda (Figura1), implementado en la función `_butter_bandpass_filter` del módulo de reducción de ruido implementado en este proyecto, para filtrar la señal y conservar únicamente las frecuencias relevantes para la identificación de especies. Sin embargo, la superposición de sonidos no deseados con las vocalizaciones de las especies objetivo sigue siendo uno de los mayores desafíos en el procesamiento automatizado de grabaciones de campo [7].

La elección de la frecuencia de corte baja del filtro Butterworth se fundamenta en el conocimiento del problema y puede estar relacionada con las características específicas de las especies de interés o con las propiedades del entorno acústico en el que se grabaron los sonidos. Esta decisión se alinea con la necesidad de eliminar las frecuencias no deseadas, como el ruido ambiental, que pueden interferir con la identificación precisa de las especies. Por otro lado, la frecuencia de corte alta del filtro está determinada por la frecuencia de Nyquist.

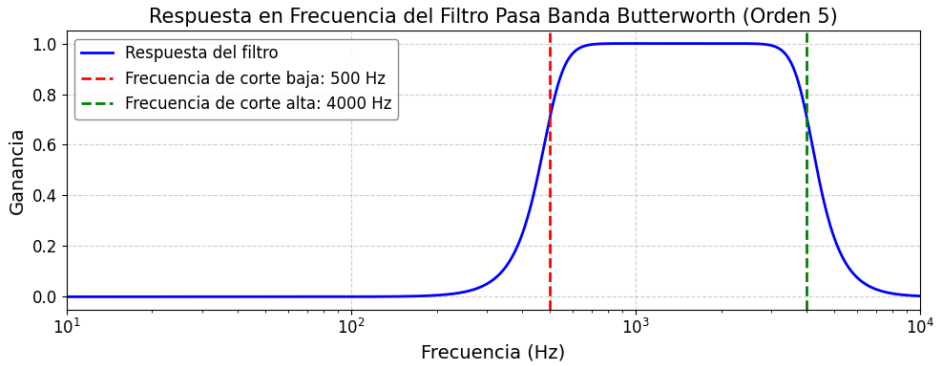


FIG. 1: Visualización de la respuesta en frecuencia de un filtro pasa banda Butterworth de orden 5, con un rango de frecuencias de 500 Hz a 4000 Hz.

IV-A4) Descomposición Wavelet: La descomposición wavelet es una técnica de análisis multiresolución que permite representar una señal en diferentes escalas y niveles de detalle. A través de esta técnica, una señal $x[k]$ se descompone en componentes de baja y alta frecuencia mediante el uso de la función madre wavelet ψ. Esta función se desplaza y escala para adaptarse a las características de la señal en distintos niveles de resolución. Además, estudios recientes han mostrado que la eliminación de ruido basado en wavelets es efectiva para separar el ruido de las vocalizaciones en grabaciones bioacústicas, destacando su utilidad en entornos con energía de baja frecuencia, como en los cantos de ballenas, ranas y aves [14]-[16]

La descomposición wavelet discreta se expresa como:

$$W_\psi[m,n] = \frac{1}{\sqrt{2^m}} \sum_{k=0}^{N-1} x[k] \cdot \psi\left(\frac{k-2^m n}{2^m}\right) \tag{3}$$

Donde $W_\psi[m,n]$ representa los coeficientes wavelet discretos en el nivel de escala m y posición n, $x[k]$ es la señal de entrada discreta en el dominio temporal, $\psi\left(\frac{k-2^m n}{2^m}\right)$ es la wavelet madre escalada y desplazada, m es el nivel de escala, que indica la cantidad de detalle en la descomposición, siendo un valor mayor de m correspondiente a una mayor escala (análisis de componentes de baja frecuencia), n es el parámetro de traslación que

determina el desplazamiento de la wavelet madre en el dominio temporal, N es la longitud de la señal, y k es el índice discreto de tiempo de la señal original.

Esta descomposición permite analizar la señal en diferentes niveles de resolución, donde los componentes de alta frecuencia corresponden a los detalles de la señal y los de baja frecuencia (aproximación) capturan su forma general [17]. Este enfoque es útil en aplicaciones como la compresión de señales, reducción de ruido, entre otros.

IV-A5) Umbralización de Coeficientes Wavelet: La umbralización de coeficientes wavelet es un paso crucial en la reducción de ruido en señales bioacústicas. Este proceso permite eliminar componentes de ruido presentes en las señales sin perder información relevante, como se observa en las Figuras 2 y 3. Los métodos de umbralización más utilizados incluyen `Visushrink`, `Sureshrink` y `Bayesshrink`, cada uno de los cuales selecciona el umbral óptimo para filtrar los coeficientes de detalle en la descomposición wavelet [16], [18]-[20]. La selección de los métodos de umbralización, como `Visushrink`, `Sureshrink` y `Bayesshrink`, es orquestada por el método `_thselect` del módulo de reducción de ruido implementado en este proyecto, asegurando una elección adaptativa del umbral en función de las características de los coeficientes de detalle wavelet [21].

En 1994, Donoho y Johnstone presentaron un enfoque clave en la reducción de ruido utilizando wavelets. Tras la descomposición wavelet, se aplica un umbral para eliminar los coeficientes asociados al ruido, este proceso se conoce como thresholding [18]. La ecuación del umbral de Donoho es utilizada en el método conocido como VisuShrink y se expresa como:

$$\tau_j = \sigma_j \sqrt{2\log(n_j)} \tag{4}$$

Donde τ_j es el umbral aplicado en el nivel de descomposición j, σ_j es la desviación estándar del nivel j, que refleja el nivel de ruido en los coeficientes wavelet, y n_j es el número de muestras de la señal o coeficientes en ese nivel.

Es posible implementar técnicas más robustas, como el uso de la Desviación Absoluta Mediana (MAD) en lugar de la desviación estándar para estimar el nivel de ruido en la señal. El MAD es menos sensible a valores atípicos en comparación con la desviación estándar, lo que lo hace más adecuado para datos con distribuciones no normales o con presencia de valores extremos. Esta técnica de estimación del ruido basada en MAD también puede aplicarse a otros tipos de umbrales, como SureShrink y BayesShrink, mejorando la robustez en la detección de coeficientes relacionados con el ruido. El umbral modificado usando MAD es:

$$\tau_j = \frac{\text{MAD}(c_j)}{0{,}6745}\sqrt{2\log(n_j)} \tag{5}$$

Donde el factor $0{,}6745$ proviene del percentil 75 de la distribución normal estándar y ajusta la MAD para estimar la desviación estándar en presencia de ruido gaussiano.

IV-A6) SureShrink: Es otra técnica propuesta por Donoho y Johnstone que se basa en minimizar el riesgo esperado mediante el uso de un criterio llamado ”Stein’s Unbiased Risk Estimate”(SURE). Este método selecciona un umbral τ_j que minimiza la pérdida cuadrática esperada. Para normalizar el ruido, se divide por sigma antes y se multiplica después del cálculo del umbral. La ecuación del umbral para SureShrink es:

$$R = \{r_i\}_{i=1,2,3,\ldots,N} = \frac{[N - 2i + (N-i)\omega_i + \sum_{k=1}^{l}\omega_k]}{N}$$

$$\tau_j = \frac{\text{MAD}}{0{,}6745}\sqrt{\omega_b}$$

Donde R es el vector de riesgo, r_i es el valor del riesgo asociado al coeficiente i, N es el número total de coeficientes, ω_i es el i-ésimo coeficiente wavelet al cuadrado, l es el límite superior para la sumatoria de los coeficientes, MAD es la desviación absoluta de la mediana de los coeficientes, τ_j es el umbral seleccionado, ω_b es el b-ésimo coeficiente wavelet al cuadrado (coeficiente correspondiente al riesgo mínimo), y $0{,}6745$ es el factor de corrección para estimar la desviación estándar de la señal ruidosa.

IV-A7) BayesShrink: Introducido por Chang, Yu y Vetterli en 2000, es un método bayesiano que selecciona un umbral que minimiza el error cuadrático medio bajo un modelo de ruido gaussiano. BayesShrink asume que el ruido wavelet sigue una distribución gaussiana blanca aditiva, El umbral óptimo en BayesShrink se calcula para minimizar el error cuadrático medio (MSE) y se define como:

$$\tau_j = \frac{\left(\frac{\text{MAD}(c_j)}{0{,}6745}\right)^2}{\sigma_s} \tag{6}$$

Donde $\frac{\text{MAD}(c_j)}{0{,}6745}$ es la estimación robusta de la desviación estándar del ruido en el nivel j, y σ_s es la desviación estándar de los coeficientes wavelet de la señal (detalles de la señal).

Los tres métodos presentados se utilizan de manera que varían en cada nivel de descomposición, lo que los hace flexibles. Este enfoque es especialmente eficaz cuando el ruido no es

constante en toda la señal, ya que ajusta dinámicamente el umbral según las características estadísticas de los coeficientes wavelet.

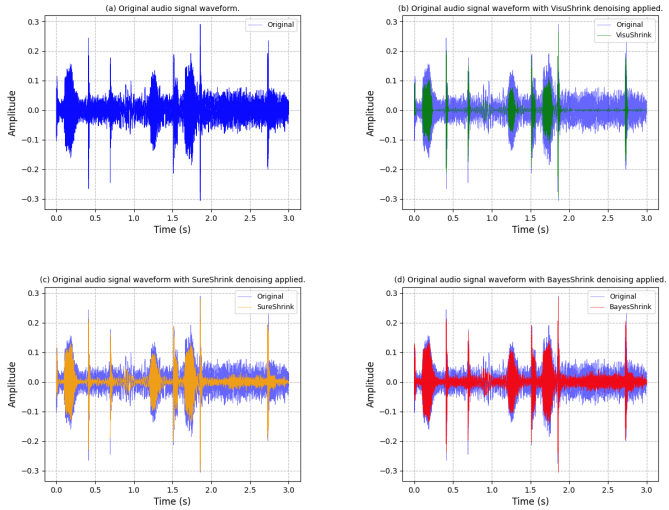


FIG. 2: Columbina squammata. Formas de onda de las señales de audio originales y denoised aplicando diferentes métodos de umbral. Se utilizó el wavelet **dmey** con nivel 9 en todos los casos: **VisuShrink**: Correlación: 0.7, Similitud coseno (Mel): 0.932, Ratio de entropía: 0.898. **SureShrink**: Correlación: 0.764, Similitud coseno (Mel): 0.944, Ratio de entropía: 0.918. **BayesShrink**: Correlación: 0.776, Similitud coseno (Mel): 0.944, Ratio de entropía: 0.928.

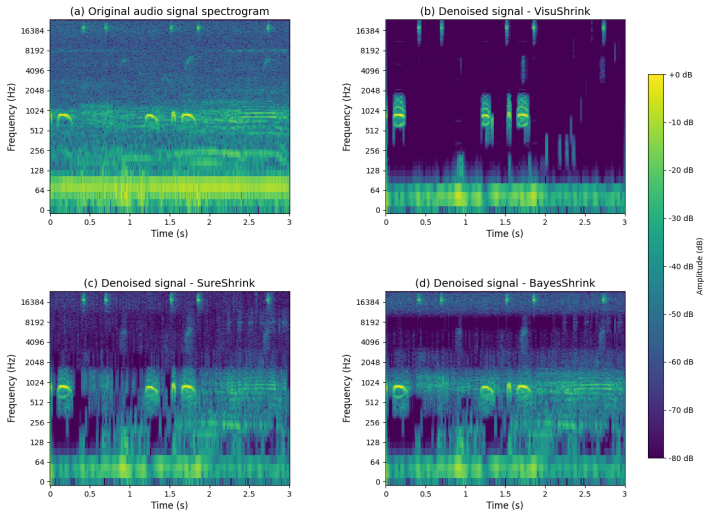


FIG. 3: Columbina squammata. Espectrogramas de las señales de audio originales y denoised aplicando diferentes métodos de umbral. Se utilizó el wavelet **dmey** con nivel 9 en todos los casos: **VisuShrink**: Correlación: 0.7, Similitud coseno (Mel): 0.932, Ratio de entropía: 0.898. **SureShrink**: Correlación: 0.764, Similitud coseno (Mel): 0.944, Ratio de entropía: 0.918. **BayesShrink**: Correlación: 0.776, Similitud coseno (Mel): 0.944, Ratio de entropía: 0.928.

IV-A8) Selección del Nivel Óptimo de Wavelet: La selección del nivel óptimo de descomposición wavelet es crucial para el proceso de eliminación de ruido. En este trabajo, la entropía de Shannon se estima sobre los coeficientes de aproximación (Figuta **fig:entropia˙metodos**), que contienen la información más representativa de la señal, para determinar el nivel óptimo de descomposición. El criterio utilizado es seleccionar el nivel en el cual la entropía comienza a aumentar respecto al nivel anterior, lo que indica que se ha alcanzado el punto donde se maximiza la retención de información estructural en la señal [22], [23]. Esta metodología ha mostrado buenos resultados sobre una base de datos de prueba de sonidos de búhos, los cuales fueron contaminados con diferentes niveles de ruido coloreado para evaluar su rendimiento [8], sin embargo, en este enfoque conservamos la ultima aproximación. Este ha sido implementado en el método `_find_optimal_wavelet_level` del script de reducción de ruido desarrollado en este proyecto.

En los casos donde no se cumple este primer criterio, se emplea un criterio alternativo basado en el ”método del codo”. Este método traza una línea recta entre los valores de entropía del primer y último nivel de descomposición y luego calcula la distancia perpendicular de cada nivel a esa línea. El nivel correspondiente a la distancia máxima se selecciona como el óptimo, ya que este representa el punto donde la disminución de la entropía deja de ser significativa. Este enfoque es particularmente útil en señales donde la entropía continúa disminuyendo de forma lenta sin mostrar un cambio claro, como ocurre en señales ruidosas o con características difusas. Al seleccionar este nivel, se asegura un equilibrio entre la complejidad de la señal y su retención de información estructural. Este criterio es viable para estos casos, ya que ofrece una selección adaptativa, evitando que el algoritmo dependa exclusivamente de un mínimo global que puede llevar a un sobreajuste, pérdida de información relevante y acumulación de error numérico innecesario.

La entropía de Shannon se define como:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log(p(x_i)) \tag{7}$$

donde X es una variable aleatoria, $p(x_i)$ es la probabilidad de ocurrencia del evento x_i y n es el número total de eventos.

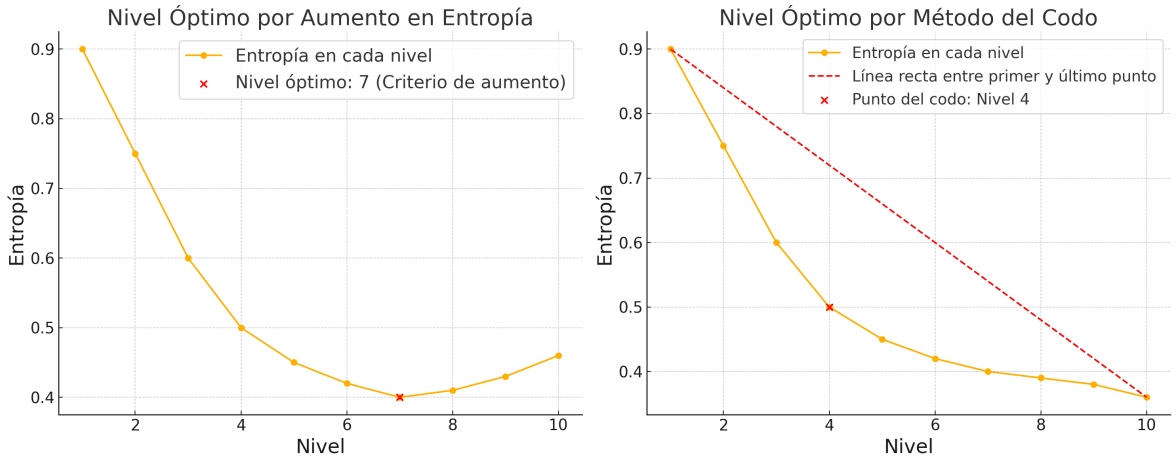


FIG. 4: Visualización comparativa de los métodos de selección del nivel óptimo de descomposición wavelet. Se muestran dos enfoques: el criterio de aumento de entropía y el método del codo. Los datos representados son sintéticos, generados para ejemplificar los métodos.

IV-A9) Correlación de Pearson: La correlación de Pearson es una métrica ampliamente utilizada en el estado del arte para evaluar la similitud entre la señal original y su versión reconstruida tras aplicar técnicas de reducción de ruido. Esta métrica cuantifica la relación lineal entre ambas señales, permitiendo medir hasta qué punto la estructura de la señal original se mantiene después del proceso de denoising, siendo aplicada en diversos estudios de procesamiento de señales acústicas [24], [25].

En el contexto del denoising basado en wavelets, la correlación de Pearson se calcula entre la señal original y la versión denoised para cada combinación de wavelet y nivel de descomposición. El proceso iterativo evalúa el rendimiento de cada combinación, siendo uno de los principales desafíos la selección de la madre wavelet más adecuada para preservar las características relevantes de la señal. La variabilidad en los resultados según la wavelet madre aplicada puede ser considerable, lo que implica que diferentes wavelets madre producirán resultados distintos dependiendo de las características específicas de la señal.

La correlación de Pearson se utiliza frecuentemente como una medida cuantitativa de la similitud entre ambas señales: cuanto mayor es el coeficiente de correlación, mayor es la preservación de la información esencial de la señal original. Sin embargo, dado que la correlación de Pearson solo captura relaciones lineales entre las señales, puede no ser completamente adecuada para evaluar señales complejas donde las relaciones no lineales entre las características de la señal original y denoised son importantes.

La ecuación de correlación de Pearson entre dos variables X e Y se expresa como:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{8}$$

donde $\text{cov}(X, Y)$ es la covarianza entre X e Y, y σ_X y σ_Y son las desviaciones estándar de X e Y, respectivamente.

IV-A10) Overfitting en wavelets: Para evitar el overfitting durante el proceso exploración de tipo de wavelets, se ha adoptado una estrategia que implica limitar la correlación de Pearson entre la señal original y la reconstruida. En este contexto, el **overfitting** se refiere a la situación en la cual el proceso de reconstrucción de la señal retiene detalles innecesarios, incluyendo el ruido que se encuentra presente en la señal original. Esto ocurre cuando la reconstrucción se ajusta demasiado a la señal original, preservando características que no son relevantes para la estructura general de la señal, sino que representan fluctuaciones ruidosas.

Un criterio específico para mitigar el overfitting es evitar que el valor absoluto de la correlación de Pearson entre la señal reconstruida y la señal original sea superior a 0.99 (umbral empírico). Si la correlación es extremadamente alta, esto sugiere que la señal reconstruida ha retenido tanto las componentes significativas como el ruido, lo cual resulta en un modelo de señal que no generaliza bien.

Este enfoque permite mantener la estructura esencial de la señal, al tiempo que asegura que la señal reconstruida no se ajuste de forma excesiva al ruido. Al evitar el overfitting, se logra una señal más representativa del fenómeno subyacente y menos afectada por fluctuaciones aleatorias o artefactos de ruido.

Aunque la correlación de Pearson ha sido útil en la evaluación de señales reconstruidas en diversos contextos, en nuestro enfoque hemos optado por añadir métricas que capturan mejor las características perceptuales y frecuenciales de las señales bioacústicas, como la similitud coseno aplicada sobre los coeficientes Mel, dado que estas complementan mejor nuestro objetivo final de preservar las características relevantes para los modelos de clasificación de especies como BirdNet y AnuraSet.

IV-A11) Similitud Coseno de los Coeficientes Mel: La similitud coseno se utiliza como una métrica para evaluar la preservación de las características acústicas esenciales, representadas por los coeficientes Mel, después de aplicar técnicas de reducción de ruido utilizando wavelets. Dado que los coeficientes Mel son una representación frecuencial diseñada para aproximar la percepción auditiva humana, es crucial que el proceso de denoising preserve estas características, de modo que la señal denoised mantenga su integridad perceptual. Esto es especialmente importante en nuestro caso, ya que los modelos de clasificación utilizados, como BirdNet y AnuraSet, emplean los espectrogramas Mel como input principal para la identificación automática de especies a partir de señales acústicas.

Para calcular esta métrica, primero se genera el espectrograma Mel de la señal original y de la señal denoised, utilizando un número fijo de bandas Mel. Posteriormente, se calcula la similitud coseno entre estos espectrogramas Mel, lo que permite evaluar qué tan cercanas están las dos señales en el espacio de las características frecuenciales. La similitud coseno entre dos vectores A y B se define como:

$$\text{Similitud Coseno}(A, B) = \frac{A \cdot B}{\|A\|\|B\|} \tag{9}$$

Donde A y B son los espectrogramas Mel de las señales original y denoised, respectivamente, y $\cdot$ denota el producto punto entre ambos vectores. El resultado de esta métrica varía entre -1 y 1, donde un valor de 1 indica que los vectores son completamente similares (preservación perfecta de las características Mel), mientras que valores más cercanos a 0 o negativos sugieren una mayor disimilitud entre las señales.

Este enfoque es particularmente útil en el contexto de la clasificación de audio basada en coeficientes Mel, ya que maximizar la similitud coseno asegura que la señal denoised mantenga las propiedades necesarias para un análisis preciso en tareas posteriores de clasificación. En nuestro experimento, seleccionamos las combinaciones de wavelet y nivel de descomposición que maximizan la similitud coseno Mel, asegurando que la señal procesada preserve los componentes perceptualmente más importantes para los modelos de clasificación como BirdNet y AnuraSet. Las funciones `correlation` y `cosine_similarity_mel` del módulo de eliminación de ruido permiten evaluar estas metricas.

IV-A12) Ratio de Entropía: Se utiliza como una métrica para evaluar la efectividad del proceso de reducción de ruido (*denoising*) en señales bioacústicas. Este ratio se calcula como la relación entre la entropía de la señal denoised y la entropía de la señal original. La entropía de Shannon, utilizada en este caso, mide el grado de desorden o complejidad en una señal. Una reducción en el Ratio de Entropía después del proceso de *denoising* sugiere que la señal ha sido simplificada, eliminando componentes ruidosos mientras se preservan las características relevantes para la clasificación bioacústica. Matemáticamente, el Ratio de Entropía se define como:

$$\text{Ratio de Entropía} = \frac{\text{Entropía de la señal denoised}}{\text{Entropía de la señal original}}$$

Esta métrica es valiosa en la evaluación de técnicas de *denoising*, ya que un valor de ratio cercano a 1 indica que la señal denoised mantiene la misma complejidad que la señal original, mientras que valores menores a 1 sugieren una reducción efectiva del ruido o complejidad de la señal.

IV-B. Modelos de aprendizaje estadístico

Los modelos de aprendizaje, como las redes neuronales, son pilares en la inteligencia artificial y el análisis de datos. Estas redes manipulan grandes volúmenes de datos para aprender y resolver tareas complejas. Su impacto abarca áreas como la medicina, la conducción autónoma, el procesamiento de lenguaje natural, la robótica, la industria manufacturera, la conservación ambiental, entre otras. Comprender su funcionamiento es fundamental para explorar su potencial, para una mayor comprensión de éstas consulte [26], [27].

IV-B1) Modelos de Clasificación de Especies: Los modelos de clasificación profunda, como ResNet152 para AnuraSet y BirdNet, utilizan arquitecturas de redes neuronales convolucionales (CNN) para extraer características discriminativas de las señales de audio. Estos modelos se entrenan utilizando espectrogramas Mel, que representan la distribución de energía de la señal en función del tiempo y la frecuencia, lo que permite capturar características relevantes para la clasificación de especies.

En este contexto, se destaca la importancia de las redes neuronales convolucionales (CNN) en tareas de visión por computadora [26], entre estas la clasificación de especies a partir de grabaciones de audio. Las CNN son modelos de aprendizaje profundo especialmente diseñados para procesar datos con una estructura de grilla, como imágenes o, en este caso, espectrogramas de audio. ResNet152 es una arquitectura CNN profunda que se ha destacado en tareas de clasificación de imágenes debido a su capacidad para aprender características jerárquicas complejas. Esta arquitectura se caracteriza por su profundidad y su capacidad para mitigar el problema de desvanecimiento del gradiente mediante el uso de conexiones residuales, lo que la hace especialmente adecuada para tareas de clasificación complejas como la identificación de especies a partir de grabaciones de audio.

La transformada de Fourier se utiliza para obtener el espectro de frecuencia de la señal de audio. Sin embargo, el espectrograma Mel, que se calcula utilizando bancos de filtros Mel, proporciona una representación más cercana a la percepción auditiva humana [7]. La ecuación que describe el flujo de todas las operaciones para calcular los coeficientes espectrales Mel (MFCC) para una trama específica de una señal de audio se puede expresar de la siguiente manera:

$$\text{MFCC}(n, i) = \text{DCT}\left(\log\left(\sum_{k=1}^{N} P_k(n) \cdot H_{i,k}\right)\right) \tag{10}$$

Donde $\text{MFCC}(n, i)$ representa el coeficiente espectral Mel (MFCC) i para la trama n de la señal de audio, $P_k(n)$ es la potencia espectral en el bin de frecuencia k de la trama n, y $H_{i,k}$ es la respuesta del filtro Mel i en la frecuencia k. La suma interna calcula la

energía ponderada en cada banda Mel, y el logaritmo se toma para estabilizar la energía y simular la percepción no lineal del oído humano. Finalmente, se aplica una Transformada Coseno Discreta (DCT) para decorrelacionar los coeficientes y representarlos de manera más compacta.

IV-B2) Metodología de evaluación: La evaluación del rendimiento del modelo en el contexto de reducción de ruido se basa en un análisis detallado de la capacidad del modelo para clasificar adecuadamente señales bioacústicas en dos escenarios distintos. Para este propósito, se observaron métricas clave como la correlación temporal, la similitud coseno en el dominio Mel, la razón de entropía y la confianza del modelo, que ofrecen una visión integral de cómo el proceso de denoising impacta la preservación de características relevantes en la señal. Estas métricas permiten cuantificar tanto la retención de las estructuras temporales y frecuenciales esenciales para la identificación de especies como la capacidad de reducir la complejidad de la señal al eliminar componentes de ruido.

IV-B2a) Ruido vs. Vocalizaciones de Especies: En este escenario, se evaluó la capacidad del modelo para discriminar entre segmentos de audio que contienen cualquier tipo de vocalización de especies y aquellos que consisten exclusivamente en ruido ambiental seleccionando la especie con mayor confianza acumulada entre las inferencias del modelo. En el conjunto de datos Parex, las muestras etiquetadas como ruido fueron seleccionadas automáticamente durante la segmentación de los intervalos temporales que no contenían vocalizaciones, utilizando criterios como una elevada energía y alta entropía en las señales, lo cual permitió crear un conjunto de datos balanceado en cuanto a muestras de audio con eventos de aves y ruido ambiental. Por otro lado, el conjunto de datos AnuraSet ya incluía etiquetas de ruido predefinidas, evitando la necesidad de un procesamiento adicional de segmentación.

IV-B2b) Vocalizaciones de Especies Frecuentes vs. Ruido: En este escenario, el análisis se centró en evaluar la capacidad del modelo para identificar y diferenciar las vocalizaciones pertenecientes a las tres especies más abundantes en el conjunto de datos frente a las vocalizaciones de otras especies, que en este contexto se consideraron como ruido. Para cada especie, se estableció una configuración en la que dicha especie se seleccionó como el objetivo principal, mientras que las vocalizaciones de las otras dos especies se asumieron como ruido. Así, en cada combinación, el modelo debe priorizar y clasificar correctamente las vocalizaciones de la especie objetivo entre las vocalizaciones de las otras especies consideradas como ruido. Este enfoque permite evaluar de forma exhaustiva la habilidad del modelo para distinguir entre las vocalizaciones más representativas de cada especie.

IV-B2c) Impacto de la Transformada Discreta Wavelet y Selección de Métodos: La metodología de evaluación se centró en analizar el impacto de la eliminación de ruido en señales bioacústicas mediante la transformada discreta wavelet, utilizando el modelo BirdNET y AnuraSet para medir la efectividad. Para este análisis, se seleccionaron todas las wavelets de la librería PyWavelets y se probaron tres métodos de umbralización (BayesShrink, SureShrink y VisuShrink), aplicados en dos escenarios de ruido: uno con solo ruido ambiental y otro con ruido y eventos bioacústicos. La métrica para la selección de las wavelets se definió como la multiplicación de la mediana de la confianza del modelo en sus predicciones por la frecuencia de selección de cada wavelet, permitiendo evaluar tanto la capacidad de cada wavelet para preservar características bioacústicas como su consistencia en ser elegida para la tarea de denoising.

IV-B3) Gráficos Radar y Pruebas de Hipótesis: Para analizar la integridad de las señales *denoised*, se emplearon gráficos radar que representan métricas clave como la correlación temporal, la similitud coseno en el dominio Mel y la razón de entropía, junto con la confianza del modelo en predicciones correctas e incorrectas. Estos gráficos proporcionan una visión detallada del efecto del denoising sobre la estructura temporal, frecuencial y la complejidad de las señales. Además, se realizó un análisis de proporciones para comparar la precisión de clasificación del modelo en señales originales y procesadas. Para evaluar diferencias en la confianza del modelo, se utilizó la prueba de Mann-Whitney U para comparar las medianas de confianza entre los audios *denoised* y los audios originales en predicciones correctas.

Se formularon dos hipótesis para esta prueba: la hipótesis nula (H_0), que sostiene que la mediana de la confianza en los audios *denoised* no es significativamente mayor que en los audios originales, y la hipótesis alternativa (H_1), que plantea que la mediana de la confianza en los audios *denoised* es mayor. Dado que se utilizó la opción `greater`, se evaluó específicamente si la confianza en los audios *denoised* supera a la de los originales. La prueba se realizó con un nivel de significancia de 0.05 para determinar si el aumento en la confianza es estadísticamente significativo.

IV-B4) Curvas ROC y Evaluación del AUC: Para estos escenarios se generaron curvas ROC (Receiver operating characteristic) y se calculó el AUC (Area Under the Curve) utilizando una metodología de *bootstrap* para obtener estimaciones robustas del AUC e intervalos de confianza para las curvas ROC. El *bootstrap* es una técnica de remuestreo con reemplazo que permite estimar la variabilidad de una métrica, en este caso el AUC, mediante la generación de múltiples subconjuntos del conjunto de prueba. Para este análisis, se realizaron 1000 iteraciones de *bootstrap*, en las cuales se generaron subconjuntos remuestreados y se calcularon las correspondientes curvas ROC para cada muestra.

Durante este proceso, las curvas ROC obtenidas en cada iteración fueron interpoladas para alinear sus valores de la tasa de falsos positivos (FPR) a un conjunto común de puntos. Esto facilitó el cálculo de la mediana de las tasas de verdaderos positivos (TPR) y aseguró una comparación consistente entre las curvas. Esta interpolación es crucial para garantizar la robustez de las métricas calculadas. Adicionalmente, se calcularon los límites de confianza del 95 % para las curvas ROC a partir de los percentiles 2.5 y 97.5 de las TPR interpoladas, proporcionando una medida de la incertidumbre asociada al modelo.

De manera similar, el AUC se estimó de forma robusta tomando la mediana de los valores obtenidos en cada iteración de *bootstrap*, lo que permitió una mejor caracterización de la capacidad del modelo para discriminar correctamente entre clases en escenarios con variabilidad inherente en los datos. La ecuación general para el cálculo del AUC es:

$$AUC = \int_0^1 TPR(FPR)\, d(FPR) \tag{11}$$

donde TPR es la tasa de verdaderos positivos (*True Positive Rate*) y FPR es la tasa de falsos positivos (*False Positive Rate*). Esta métrica cuantifica la capacidad del modelo para discriminar entre clases a diferentes umbrales de decisión, proporcionando un valor agregado cuando se compara el AUC de señales originales frente a señales *denoised*.

IV-B5) Pruebas de Hipótesis (AUC): Adicionalmente, para validar estadísticamente la diferencia en el rendimiento entre los modelos entrenados con señales originales y señales *denoised*, se aplicó la prueba de Wilcoxon (`greater`) para muestras pareadas. Esta prueba no paramétrica es especialmente adecuada en este contexto, ya que no requiere asumir una distribución normal de las diferencias entre las muestras, lo cual es relevante dado el posible comportamiento asimétrico de los valores de AUC. La prueba de Wilcoxon se utilizó para comparar las medianas de las distribuciones de AUC obtenidas a partir de los conjuntos de señales ruidosas (*dirty*) y *denoised*, con el objetivo de determinar si la mediana de los AUC para las señales *denoised* era significativamente mayor que la de las señales *dirty*. El análisis se realizó con un nivel de significancia del 95 %. El resultado de la prueba se interpretó de la siguiente manera:

Si $p < 0{,}05$, se rechaza la hipótesis nula, indicando que los AUC en el caso *denoised* son significativamente mayores que los obtenidos con señales ruidosas. Esto confirma que el proceso de denoising mejora significativamente el rendimiento del modelo. Si $p \geq 0{,}05$, no se puede rechazar la hipótesis nula, lo que sugiere que no se encontró una diferencia significativa entre ambos casos.

IV-B6) Distancia de Jaccard: Además, la distancia de Jaccard es una métrica utilizada para medir la disimilitud entre dos conjuntos, lo cual resulta particularmente útil en problemas de clasificación multietiqueta. En este contexto, se define como la proporción de elementos diferentes entre las etiquetas predichas y las etiquetas reales de una instancia dada. Cuanto mayor es la distancia de Jaccard, menor es la coincidencia entre los conjuntos de etiquetas, lo que indica una menor precisión en la clasificación.

La distancia de Jaccard es especialmente relevante en este problema de clasificación de señales bioacústicas, ya que permite evaluar el impacto de la eliminación de ruido (*denoising*) sobre la coincidencia de etiquetas. Al comparar los conjuntos de etiquetas obtenidos con y sin *denoising*, esta métrica ayuda a identificar si el proceso de reducción de ruido está afectando la precisión de la clasificación al eliminar o atenuar características importantes.

La distancia de Jaccard entre dos conjuntos A y B se define como:

$$J(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|} \tag{12}$$

donde $|A \cap B|$ representa el número de elementos comunes entre los conjuntos y $|A \cup B|$ es el número total de elementos en la unión de ambos conjuntos. En este caso, A representa las etiquetas predichas y B las etiquetas reales. Una distancia de Jaccard cercana a 0 indica una alta coincidencia entre los conjuntos, mientras que una distancia más cercana a 1 indica una menor coincidencia.

IV-B7) Métrica utilizada para evaluar la efectividad de cada wavelet: La métrica utilizada para evaluar la efectividad de cada wavelet se define como:

$$\text{Métrica} = \text{Mediana}(\text{Confianza}) \times \text{Frecuencia de Selección}, \tag{13}$$

donde Mediana(Confianza) representa la mediana de la confianza asignada por el modelo BirdNET a las predicciones de la señal procesada con cada wavelet, y Frecuencia de Selección es el número de veces que cada wavelet fue seleccionada como óptima en el proceso de eliminación de ruido. Esta métrica permite capturar tanto la efectividad en la preservación de la señal de interés como la consistencia de la wavelet en su selección como la mejor opción para la tarea.

V. Metodología

Para abordar el problema de evaluación del impacto de las técnicas de reducción de ruido basadas en wavelets en el procesamiento de señales bioacústicas y su influencia en el rendimiento de modelos de referencia para la clasificación de datos, se implementará una metodología que sigue el Modelo de Proceso de Minería de Datos CRISP-DM (Cross Industry Standard Process for Data Mining), adaptado al contexto del procesamiento de señales bioacústicas y el aprendizaje automático.

V-A. Comprensión del negocio

En el marco de este proyecto, se subraya la importancia del monitoreo de poblaciones biológicas a través del análisis de señales acústicas, que posibilita la clasificación automatizada de especies sin la necesidad de contacto directo. Se destaca la influencia significativa de la elección entre micrófonos direccionales u omnidireccionales en la captura de datos y en la aplicación posterior de técnicas de procesamiento de señales. Los micrófonos direccionales, al concentrar la señal de interés, pueden presentar desafíos específicos en la reducción de ruido, mientras que los omnidireccionales capturan tanto la señal de interés como el ruido ambiental, planteando consideraciones adicionales en el procesamiento de datos. Por ende, comprender en profundidad las implicaciones de cada tipo de micrófono es crucial para guiar el desarrollo de técnicas de procesamiento de señales que maximicen la calidad de los datos y, por ende, la eficacia de los modelos de clasificación utilizados en el monitoreo de poblaciones biológicas.

V-B. Comprensión de los datos

Los datos utilizados en este estudio consisten en grabaciones de señales bioacústicas obtenidas mediante dispositivos especializados en entornos naturales. Estos datos fueron proporcionados por el Instituto de Investigación de Recursos Biológicos Alexander von Humboldt. En particular, los datos del conjunto *Parex* no han sido publicados previamente, lo que garantiza su autenticidad y originalidad. La recolección se realizó de manera pasiva con el propósito de monitorear la diversidad biológica en diferentes hábitats.

Por otro lado, el conjunto de datos *AnuraSet* es un benchmark público actualmente disponible, diseñado para la identificación bioacústica de anuros neotropicales. Este conjunto de datos incluye grabaciones anotadas por expertos y puede ser accedido en soundclim.github.io/anuraweb.

Las grabaciones presentan contaminación por ruido ambiental, lo que dificulta su análisis y la extracción de características relevantes para la identificación de especies. Es importante destacar que las señales bioacústicas de aves y anfibios presentan diferencias significativas

en términos de sus características acústicas y de comportamiento, lo que hace esencial la implementación de técnicas de reducción de ruido para mejorar la calidad del procesamiento.

V-B1) Parex: El conjunto de datos *Parex*, utilizado en este estudio, comprende señales bioacústicas recolectadas mediante dispositivos especializados para el monitoreo de la biodiversidad en el Magdalena Medio. Las grabaciones se obtuvieron siguiendo un muestreo sistemático en el que se seleccionaron segmentos de audio de 10 segundos cada 30 minutos, a lo largo de 12 o 13 días. Este procedimiento permitió la recolección de 48 segmentos diarios, equivalentes a 8 minutos de audio por día. Cada entrada en el conjunto de datos está representada por las siguientes variables:

- **Begin Time (s)**: Tiempo de inicio de la señal en segundos.
- **End Time (s)**: Tiempo de finalización de la señal en segundos.
- **Low Freq (Hz)**: Frecuencia baja de la señal en Hertz.
- **High Freq (Hz)**: Frecuencia alta de la señal en Hertz.
- **TIPO**: Tipo de señal, en este caso, se refiere a señales bioacústicas de aves.
- **ID**: Identificador de la señal.
- **DET**: Detalle específico de la señal (Vocalizaciones de aves).
- **Especie**: Nombre científico de la especie relacionado con el detalle específico de la señal.
- **fname**: Nombre del archivo que contiene la señal.

V-B2) AnuraSet: El conjunto de datos *AnuraSet* comprende grabaciones bioacústicas recolectadas mediante dispositivos especializados para el monitoreo de anuros neotropicales. Este conjunto de datos forma parte de un programa de monitoreo pasivo a nivel nacional en Brasil y se basa en 1.612 grabaciones de audio de un minuto, anotadas por expertos y distribuidas en cuatro sitios estratégicos dentro de los biomas del Cerrado y la Mata Atlántica.

Los datos de *AnuraSet*, incluyendo las grabaciones originales, la configuración experimental y un modelo de referencia para clasificación, están disponibles de forma abierta en el siguiente enlace: soundclim.github.io/anuraweb.

- **sample_name**: Nombre de la muestra.
- **fname**: Nombre del archivo que contiene la muestra.
- **min_t**: Tiempo mínimo de la muestra en segundos.
- **max_t**: Tiempo máximo de la muestra en segundos.

- **site**: Sitio donde se grabó la muestra.
- **date**: Fecha en que se grabó la muestra.
- **species_number**: Número de especies presentes en la muestra.
- **subset**: Subconjunto al que pertenece la muestra.
- **SPHSUR**: Especie asociada al detalle específico de la señal indicado en one hot encoding.
- **BOABIS**: ...
- **SCIPER**: ...
- **...**: (continúa para otras especies presentes)

V-C. inspección visual y auditiva de los datos

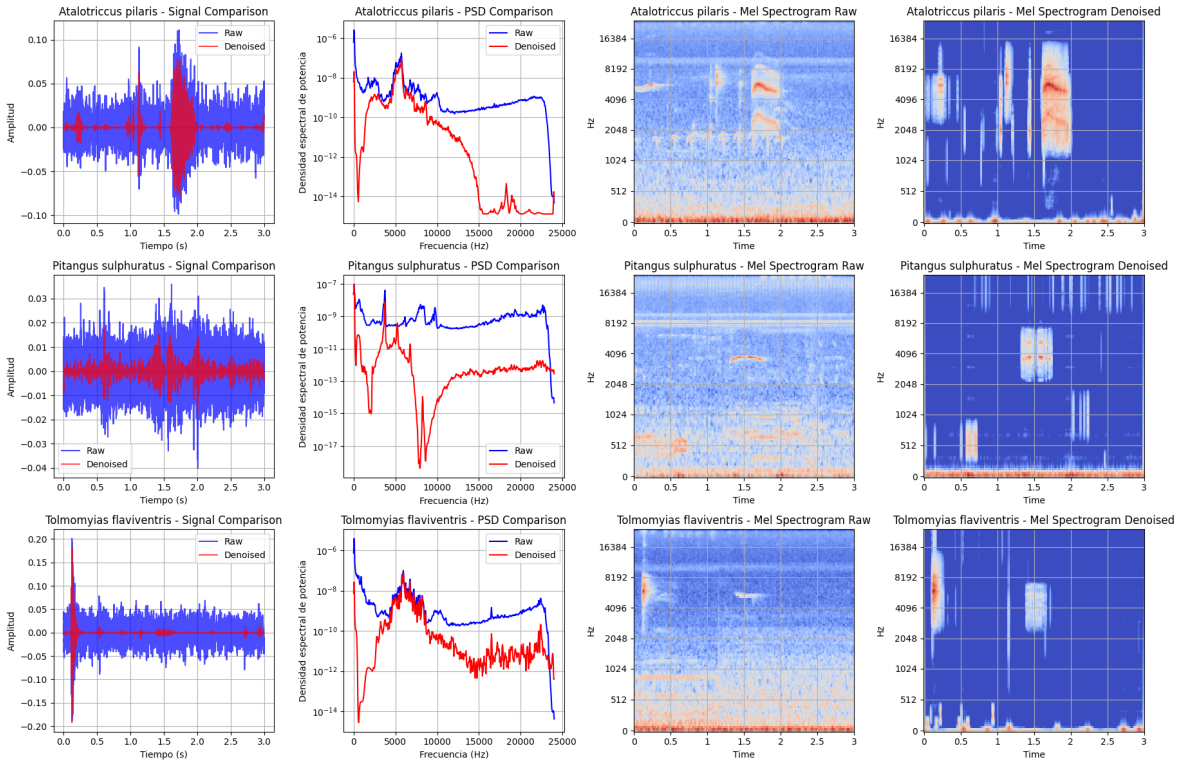


FIG. 5: Comparación de las señales bioacústicas antes y después de la reducción de ruido utilizando el método VisuShrink. Se presentan la forma de onda en el dominio del tiempo, la Densidad Espectral de Potencia (PSD) y los espectrogramas Mel para diferentes especies, permitiendo analizar el comportamiento del filtrado y la preservación de características acústicas clave.

Para comprender mejor los datos, se llevó a cabo un análisis detallado de las diferencias entre especies y se evaluó, tanto visual como auditivamente, el impacto del método de reducción de ruido VisuShrink. En primer lugar, se representaron las señales en el dominio del tiempo mediante la comparación de la forma de onda original (Raw) y la procesada (Denoised), lo que permitió observar cómo el ruido afectaba la estructura de las vocaliza-

ciones y cómo el filtrado preservaba los eventos bioacústicos de interés. Luego, se analizó la Densidad Espectral de Potencia (PSD), con el objetivo de evaluar cómo se redistribuía la energía espectral tras la eliminación del ruido, asegurando que las frecuencias clave de las especies no fueran alteradas de manera significativa.

Además, se emplearon espectrogramas Mel para visualizar la estructura tiempo-frecuencia de las vocalizaciones antes y después del procesamiento. Esto permitió no solo identificar las características acústicas específicas de cada especie, sino también evaluar la efectividad del filtrado al resaltar los eventos sonoros relevantes mientras atenuaba el ruido de fondo. Al aplicar este análisis a diferentes especies, se obtuvo una visión más clara sobre la variabilidad en las señales bioacústicas y sobre el comportamiento del método VisuShrink en distintos escenarios.

V-D. Preparación de los datos

Dado que el Instituto Humboldt ha proporcionado las muestras de audio junto con su metadata previamente etiquetada y revisada por expertos, la etapa de preparación de datos se enfocará en la mitigación del ruido presente en las señales.

Se aplicarán técnicas de reducción de ruido basadas en transformadas Wavelet para atenuar el ruido ambiental presente en las señales acústicas. Estas técnicas permitirán preservar la información esencial de las especies biológicas de interés mientras se reduce la interferencia del entorno circundante.

Cabe destacar que todo el *pipeline* de procesamiento de datos, desde la preparación hasta la evaluación de los modelos y datos, ha sido implementado en el lenguaje de programación Python, aprovechando sus bibliotecas especializadas para el análisis y procesamiento de señales.

V-E. Modelado

En esta fase, se evaluará el desempeño de los modelos de referencia previamente entrenados en las muestras de datos procesadas y no procesadas con la eliminación de ruido mediante el método `denoise_signal`. Este método, según su parametrización, coordina la aplicación de técnicas como `_find_optimal_wavelet_level` y `_thselect`, previamente mencionadas.

El objetivo de esta evaluación es determinar el impacto de las técnicas de reducción de ruido basadas en transformadas Wavelet sobre el rendimiento de los modelos de clasificación de especies biológicas. Para ello, se emplearán modelos de referencia previamente entrenados, como BirdNET y AnuraSet, diseñados específicamente para la clasificación de especies biológicas a partir de señales acústicas.

V-F. Evaluación

Se evaluará el desempeño de los modelos de referencia utilizando el conjunto de datos procesado utilizando los metodos, generando curvas ROC y calculando el área bajo la curva (AUC). Se empleará la técnica de bootstrap para obtener estimaciones robustas del AUC e intervalos de confianza en la curva ROC, evaluando así la capacidad de los modelos en condiciones de reducción de ruido.

Los resultados obtenidos se compararán entre el desempeño del modelo con los datos originales y con los datos donde se aplicaron técnicas de eliminación de ruido, para identificar posibles mejoras o deterioros en su rendimiento.

V-G. Despliegue

Como parte de la conclusión de este proyecto de investigación, se establecerá un repositorio que incluirá los resultados del modelo, acompañado de un informe detallado que explique los hallazgos y conclusiones obtenidos [28]. Adicionalmente, se planea publicar una librería desarrollada con Poetry para facilitar la implementación de los métodos propuestos y poner a disposición una aplicación interactiva basada en Gradio, diseñada para permitir la experimentación con los modelos de manera sencilla. Dado que estos modelos no están destinados a un entorno de uso continuo, no se realizará un despliegue formal. El enfoque principal será proporcionar documentación clara y precisa, de manera que la comunidad científica interesada en el análisis de señales acústicas y la conservación de la biodiversidad pueda replicar y explorar los resultados presentados. Tal como se detalla en el Anexo sobre Solicitudes bajo Confidencialidad, este proceso asegura el respeto y cumplimiento de los derechos de autor.

VI. Resultados y discusión

Los resultados se presentan en dos visualizaciones principales:

- **Gráficos Radar**: Los gráficos radar muestran las métricas clave de evaluación del modelo para los casos de predicción correcta e incorrecta, incluyendo correlación temporal, similitud coseno en el dominio Mel, razón de entropía y confianza del modelo. Los intervalos de confianza del 95 % se representan para cada métrica, junto con una referencia de confianza en el caso de audio sin denoising.

- **Gráfico de Barras de Proporciones**: Este gráfico ilustra la proporción de predicciones correctas y erróneas para los tipos de audio "Audio Original" y "Audio Denoised", permitiendo una comparación directa del efecto del denoising en la precisión de clasificación del modelo.

Aves

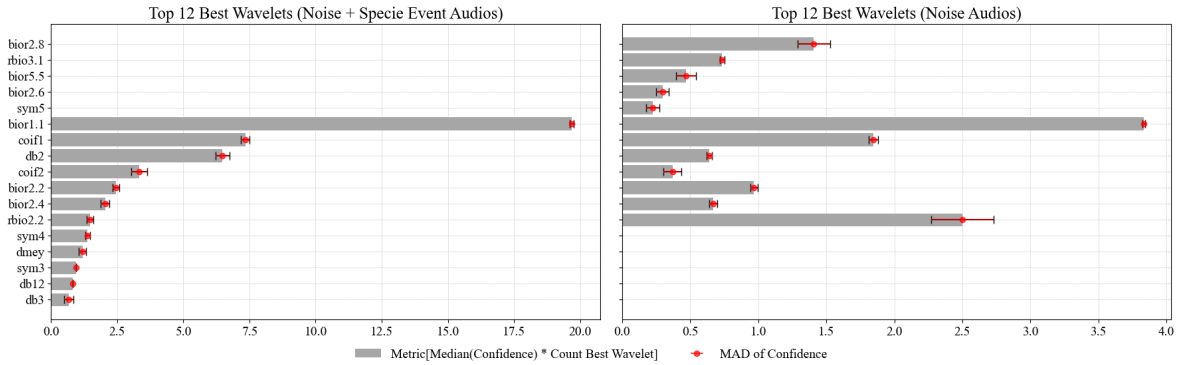


FIG. 6: Top 12 wavelets en los escenarios de "Noise + Specie Event Audios"(izquierda) y "Noise Audios"(derecha). El valor de la métrica se calcula como el producto entre la mediana de la confianza asignada por el modelo BirdNET y el número de veces que cada wavelet fue seleccionada como la mejor en el proceso de eliminación de ruido. Las barras incluyen errores calculados como MAD (Median Absolute Deviation) para la confianza.

El gráfico de la Figura 6 presenta las *top* 12 wavelets seleccionadas en dos escenarios experimentales utilizando la ecuación 13: *Noise + Specie Event Audios*, que incluye señales con ruido ambiental combinado con eventos de cantos de aves, y *Noise Audios*, correspondiente a señales compuestas únicamente por ruido ambiental. En el caso de *Noise + Specie Event Audios*, las wavelets destacadas muestran una alta capacidad para preservar características bioacústicas críticas, esenciales para la identificación y clasificación de especies en entornos acústicamente complejos. Por el contrario, las wavelets seleccionadas en *Noise Audios* optimizan la atenuación de ruido en ausencia de eventos de interés, lo que resulta efectivo

para una eliminación general de ruido sin priorizar la conservación de detalles específicos de los cantos de aves.

Existen varias estrategias para seleccionar el conjunto óptimo de wavelets: la unión de ambas selecciones incluiría wavelets con efectividad en al menos uno de los contextos (*Noise + Specie Event Audios* y *Noise Audios*), proporcionando versatilidad, aunque sin garantizar la preservación de los eventos bioacústicos. La selección de wavelets exclusivas del conjunto *Noise Audios* priorizaría la reducción de ruido en general, lo cual podría conllevar la pérdida de información relevante en presencia de señales bioacústicas. Por otro lado, centrarse en las wavelets exclusivas de *Noise + Specie Event Audios*, como *dmey*, *sym3*, *db12*, *db3* y *sym4*, permite enfocarse en aquellas transformadas que son especialmente eficaces en la separación de ruido mientras preservan las características acústicas de interés en el contexto de eventos de aves.

Optar por las wavelets exclusivas de *Noise + Specie Event Audios* constituye una estrategia óptima para maximizar la conservación de detalles bioacústicos específicos al tiempo que se reduce el ruido ambiental. Estas wavelets demuestran una especialización en la preservación de señales relevantes, manteniendo la integridad de los eventos de interés en entornos ruidosos. Esta elección es clave en estudios bioacústicos, donde se requiere un equilibrio entre la reducción de ruido y la preservación de las características de la señal, mejorando así la precisión y fiabilidad en la identificación y análisis de especies en contextos de monitoreo ambiental.

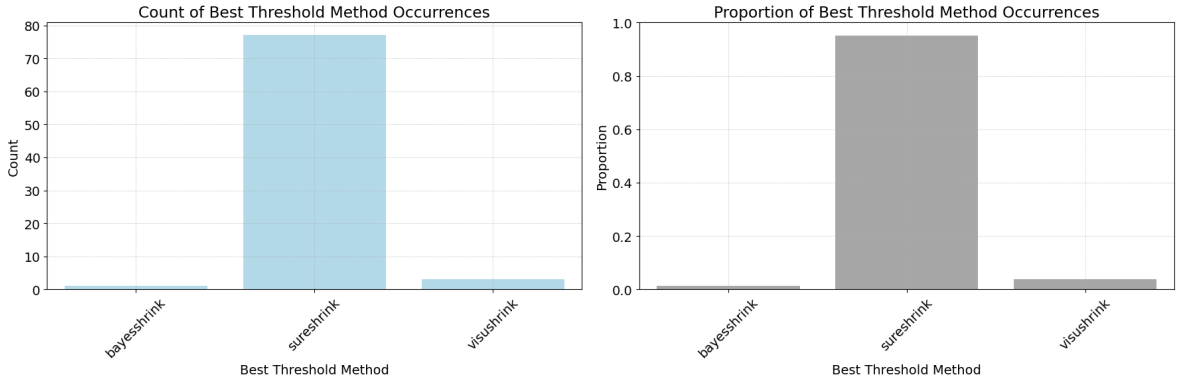


FIG. 7: Frecuencia y proporción de selección de cada método de umbralización (bayesshrink, sureshrink, visushrink) en el proceso de eliminación de ruido.

El gráfico de la Figura 7 muestra la frecuencia de selección de los métodos de umbralización *bayesshrink*, *sureshrink* y *visushrink* en el proceso de eliminación de ruido. En el panel izquierdo, se observa que el método *sureshrink* es claramente predominante, con un número de ocurrencias notablemente superior en comparación con los otros métodos, cuya frecuencia es casi despreciable. Esta tendencia se confirma en el panel derecho, donde la proporción de selección de *sureshrink* se aproxima a la totalidad, alcanzando casi el 100 %. Este

resultado indica que *sureshrink* es el método más eficaz en la configuración experimental actual, destacándose en su capacidad para preservar características bioacústicas mientras reduce el ruido. Con el fin de optimizar el costo computacional, se decide utilizar las wavelets seleccionadas previamente del conjunto *Noise + Specie Event Audios* (*dmey*, *sym3*, *db12*, *db3* y *sym4*) en combinación con el método de umbralización *sureshrink* para el análisis subsecuente, garantizando así la eficiencia y la conservación de los detalles acústicos esenciales para la identificación de especies en ambientes ruidosos.

VI-A. Ruido vs. Vocalizaciones de Especies

Este estudio analiza el experimento *Ruido vs. Vocalizaciones de Especies*, en el cual se evaluó la capacidad de un modelo de clasificación bioacústica para discriminar entre segmentos de audio con vocalizaciones de aves y aquellos que contienen exclusivamente ruido ambiental. En este contexto, se usaron los datos de Parex, donde las muestras etiquetadas como ruido fueron seleccionadas automáticamente durante la segmentación de los intervalos sin vocalizaciones, utilizando criterios como energía elevada y alta entropía. Esto permitió crear un conjunto de datos balanceado en cuanto a eventos de vocalización y ruido ambiental.

VI-A1) Gráficos Radar: Evaluación de Métricas de Denoising: En el gráfico radar izquierdo de la Figura 8 correspondiente a las predicciones incorrectas, observamos que la **correlación temporal** muestra un valor mediano de aproximadamente $0{,}714$ mientras que la **similitud coseno en el dominio Mel** alcanza $0{,}784$. Estos valores indican que, cuando el modelo no clasifica correctamente, el proceso de denoising no logra preservar de manera efectiva la estructura temporal y frecuencial de las vocalizaciones, sugiriendo que en estos casos el ruido ambiental interfiere con las características relevantes para la clasificación. La **confianza del modelo** en estos casos es baja, con un valor mediano de aproximadamente 0.59, reflejando una alta incertidumbre en las predicciones cuando la señal aún contiene componentes ruidosos significativos. Por otro lado, la **razón de entropía** presenta un valor de 0.921, lo cual indica una reducción en la cantidad de información en la señal denoised en comparación con la señal original. Una disminución en este valor sugiere que el proceso de denoising ha logrado simplificar la señal en cierta medida.

En contraste, el gráfico radar de las predicciones correctas, muestra una mejora considerable en las métricas clave. La **correlación temporal** aumenta a un valor mediano de 0.841, y la **similitud coseno en el dominio Mel** alcanza 0.907, lo que indica que el modelo retiene mejor las características temporales y frecuenciales cuando se aplican técnicas de denoising y las vocalizaciones están más claramente discernibles. La **confianza del modelo**

es significativamente mayor en estos casos, alcanzando un valor mediano de 0.643, lo cual sugiere que el proceso de reducción de ruido facilita una clasificación más precisa y confiable. Comparado con la señal original sin procesamiento de reducción de ruido, en la cual la confianza tenía un valor de 0.481, el aumento en la confianza evidencia el beneficio del denoising en la seguridad de las predicciones del modelo. La **razón de entropía** es de 0.948, reflejando una reducción en la complejidad de la señal, alineándose con el objetivo de la reducción de ruido de minimizar la interferencia ambiental manteniendo la integridad de la vocalización.

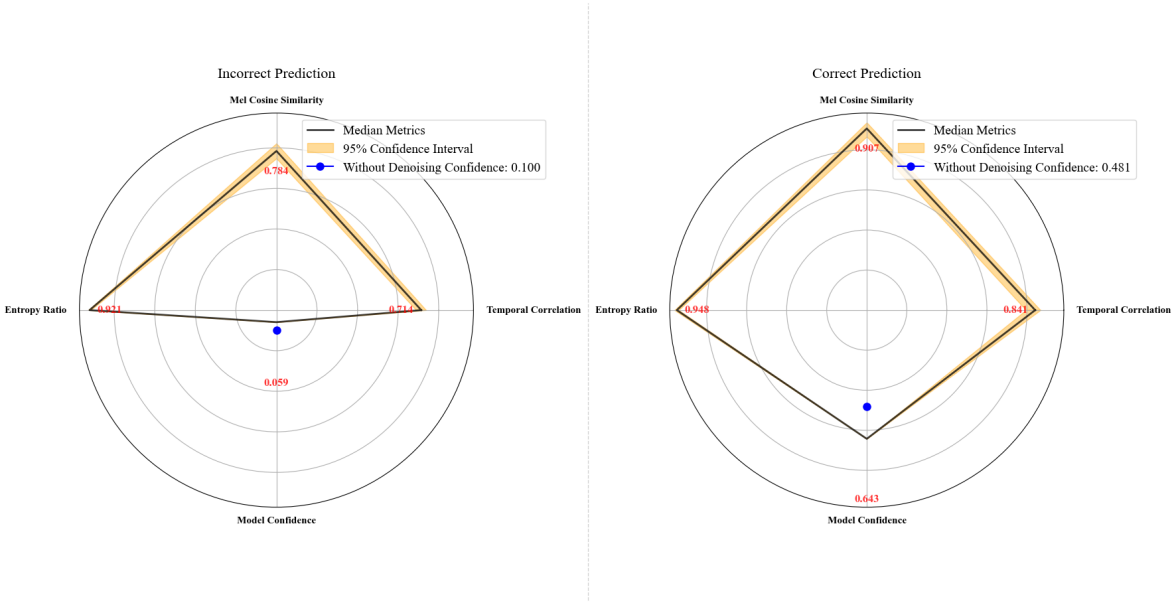


FIG. 8: Gráficos radar de métricas de evaluación para casos con predicción correcta (derecha) e incorrecta (izquierda). Los gráficos incluyen los intervalos de confianza del 95 % para cada métrica y una referencia de confianza en el caso sin denoising.

Los resultados de la prueba de Mann-Whitney U para los niveles de confianza fueron los siguientes:

- **Estadístico U**: 8148.0
- **Valor p**: 0.0887

Dado que el valor p ($p = 0{,}0887$) es mayor que el nivel de significancia comúnmente utilizado de 0.05, no hay suficiente evidencia estadística para rechazar la hipótesis nula en este caso. Esto indica que la mejora en la confianza del modelo entre los audios *denoised* y los originales no es significativa al 5 % de significancia. En otras palabras, aunque se observó un aumento en la confianza con el proceso de *denoising* en las predicciones correctas, este aumento no es estadísticamente significativo en comparación con la muestra de audios originales sin *denoising*.

VI-A2) Gráfico de Barras de Proporciones:: El gráfico de barras de proporciones (Figura 9) muestra que el "Audio Denoised" presenta un **13.8 % de predicciones correctas** y un **86.2 % de predicciones incorrectas**, mientras que el "Audio Original" tiene un **20.0 % de predicciones correctas** y un **80.0 % de predicciones incorrectas**. Aunque el denoising mejora la confianza del modelo y ayuda a retener características relevantes de las vocalizaciones, la proporción de especies identificadas correctamente es menor en el caso denoised. Esto sugiere que, si bien la eliminación de ruido reduce la interferencia y simplifica la señal, podría estar suprimiendo detalles acústicos sutiles necesarios para una identificación precisa de algunas especies o muestras.

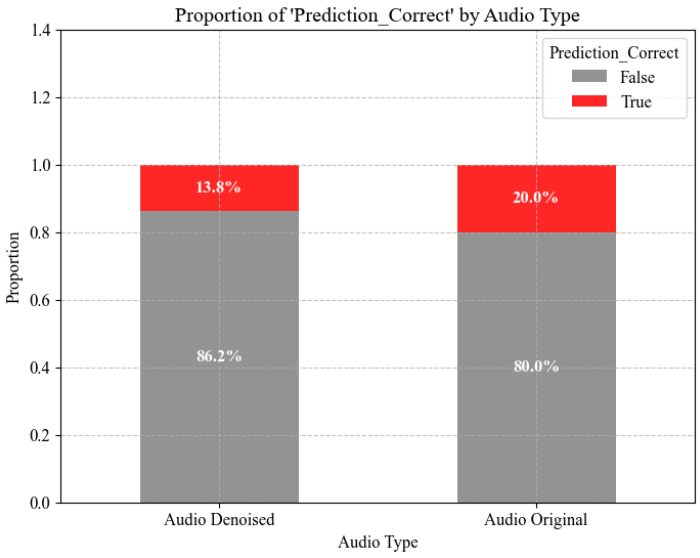


FIG. 9: Proporción de predicciones correctas y erróneas para los tipos de audio "Audio Original" y "Audio Denoised" (679 muestras).

VI-A3) Análisis de la Distancia de Jaccard: La distancia de Jaccard se calculó para comparar la precisión en la clasificación multietiqueta entre audios con y sin eliminación de ruido. Los resultados muestran que la distancia de Jaccard para el caso con eliminación de ruido es de 0.9292, mientras que para el caso sin eliminación de ruido es de 0.8777. Este incremento en la distancia de Jaccard sugiere una menor coincidencia entre las etiquetas predichas y las etiquetas reales en el caso con denoising en comparación con el caso sin procesamiento de reducción de ruido. Esto indica que, aunque la eliminación de ruido puede reducir la complejidad de la señal y mejorar ciertos aspectos de la clasificación, también podría estar eliminando o atenuando características importantes necesarias para la identificación precisa de etiquetas específicas.

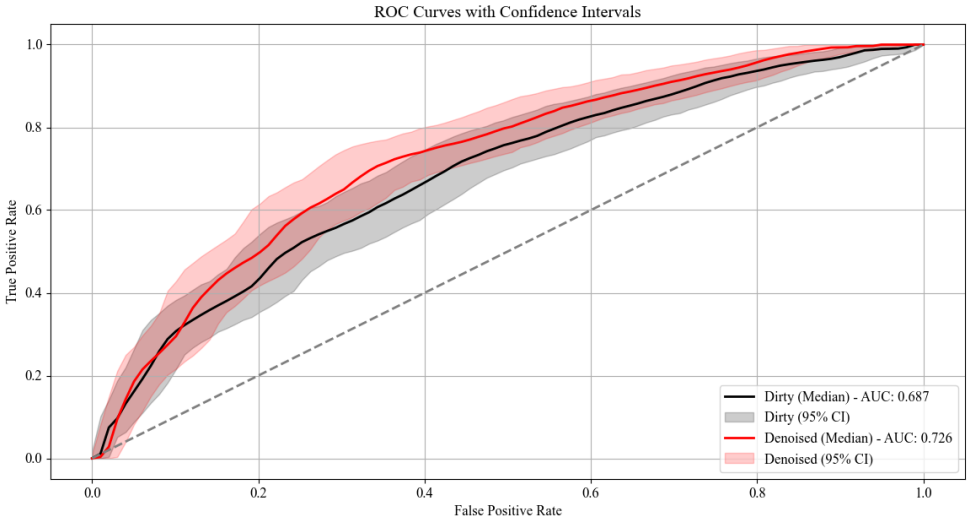


FIG. 10: Curvas ROC con intervalos de confianza para los casos Dirty (Audio Original) y Denoised (Eliminación de Ruido). La curva en negro representa el rendimiento del modelo con señales originales, con un AUC de 0.687, y su intervalo de confianza al 95 % sombreado en gris. La curva en rojo muestra el rendimiento del modelo con señales denoised, alcanzando un AUC de 0.726, con su intervalo de confianza al 95 % sombreado en rojo. La mejora en el AUC sugiere que el denoising facilita la clasificación correcta al reducir la interferencia del ruido ambiental.

VI-A4) Análisis de la Curva ROC: La figura 10 muestra curvas ROC con intervalos de confianza para dos condiciones: señales originales (Dirty) y señales denoised (Denoised). La curva en negro representa el rendimiento del modelo con señales originales, con un área bajo la curva (AUC) de 0.687 y su intervalo de confianza del 95 % sombreado en gris. La curva en rojo representa el rendimiento con señales denoised, alcanzando un AUC de 0.726 y con su intervalo de confianza sombreado en rojo. La mejora en el AUC para el caso de denoising indica que el modelo clasifica de manera más efectiva en presencia de reducción de ruido, lo que sugiere que el denoising facilita la identificación correcta al reducir la interferencia de ruido ambiental. Esta diferencia en el AUC es especialmente notable en el rango de falsos positivos bajos a intermedios, donde la curva ROC para el caso denoised muestra una mayor tasa de verdaderos positivos. Esto implica que el denoising permite al modelo detectar mejor las vocalizaciones de especies mientras minimiza las clasificaciones erróneas de ruido ambiental como señales de interés, contribuyendo así a un rendimiento general mejorado en la clasificación.

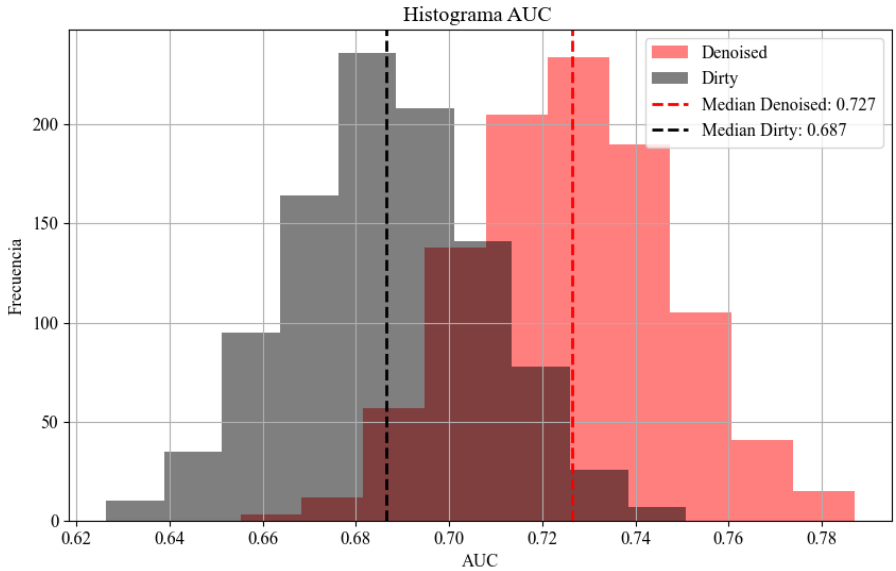


FIG. 11: Histogramas de AUC obtenidos mediante bootstrapping para los modelos denoised y dirty (Audios sin procesar). Las líneas punteadas representan las medianas para cada caso: 0.727 para denoised y 0.687 para dirty.

La Figura 11 presenta histogramas de las distribuciones del AUC obtenidas mediante bootstrapping para dos condiciones de procesamiento: señales denoised (rojo) y señales dirty (negro). La mediana del AUC en el caso denoised es de 0.727, mientras que en el caso dirty es de 0.687, lo que indica una mejora notable en el rendimiento del modelo al aplicar técnicas de reducción de ruido. La separación visible entre las dos distribuciones y la concentración de valores de AUC más altos en el modelo denoised sugieren que este logra una clasificación más efectiva al reducir la interferencia de ruido. Además, los resultados de la prueba de Wilcoxon, con un estadístico de 485186 y un p-valor de $3{,}806 \times 10^{-146}$, permiten rechazar la hipótesis nula al nivel de significancia de 0.05. Esto proporciona una fuerte evidencia estadística para concluir que los AUC de las señales denoised son significativamente mayores que los de las señales dirty, lo que implica que el modelo en la condición denoised ofrece una calidad en la diferenciación entre casos donde hay alguna especie y aquellos en donde solo hay ruido ambiental.

VI-B. Vocalizaciones de Especies Frecuentes vs. Ruido

En esta sección se analizan los resultados específicos para cada una de las especies más frecuentes en el conjunto de datos Parex: *Atalotriccus pilaris*, *Pitangus sulphuratus* y *Tolmomyias flaviventris*. A continuación, se presentan los resultados y observaciones para cada una de estas especies en función de las métricas evaluadas tras la aplicación de denoising.

VI-B1) Resultados para Atalotriccus pilaris: Para los casos de predicción incorrecta, el valor mediano de la correlación temporal es de aproximadamente $0{,}775$, indicando que el modelo pierde cierta estructura temporal relevante al aplicar denoising en las vocalizaciones de esta especie. La similitud coseno en el dominio Mel alcanza un valor de $0{,}877$, lo que muestra una preservación parcial de la estructura frecuencial, aunque no lo suficiente para evitar clasificaciones erróneas. La razón de entropía, con un valor de $0{,}942$, sugiere que el proceso de denoising reduce la complejidad de la señal, pero también podría estar eliminando detalles cruciales para la identificación precisa de la especie. La confianza del modelo es baja en estos casos, con una mediana de $0{,}0$ ya que estos son casos en donde no se detecto la especie correcta.

En contraste, los casos de predicción correcta muestran una mejora en la correlación temporal, con un valor mediano de $0{,}858$, indicando una mayor retención de las características temporales esenciales de la vocalización de *Atalotriccus pilaris*. La similitud coseno en el dominio Mel alcanza $0{,}903$, lo que demuestra que la técnica de denoising conserva mejor la estructura frecuencial en estos casos. La razón de entropía se mantiene en $0{,}964$, reflejando una complejidad controlada que facilita la clasificación correcta. La confianza del modelo en estos casos tiene un valor mediano de $0{,}649$ siendo esta inferior a la referencia donde no se aplico el procesamiento, siendo esta de $0{,}679$.

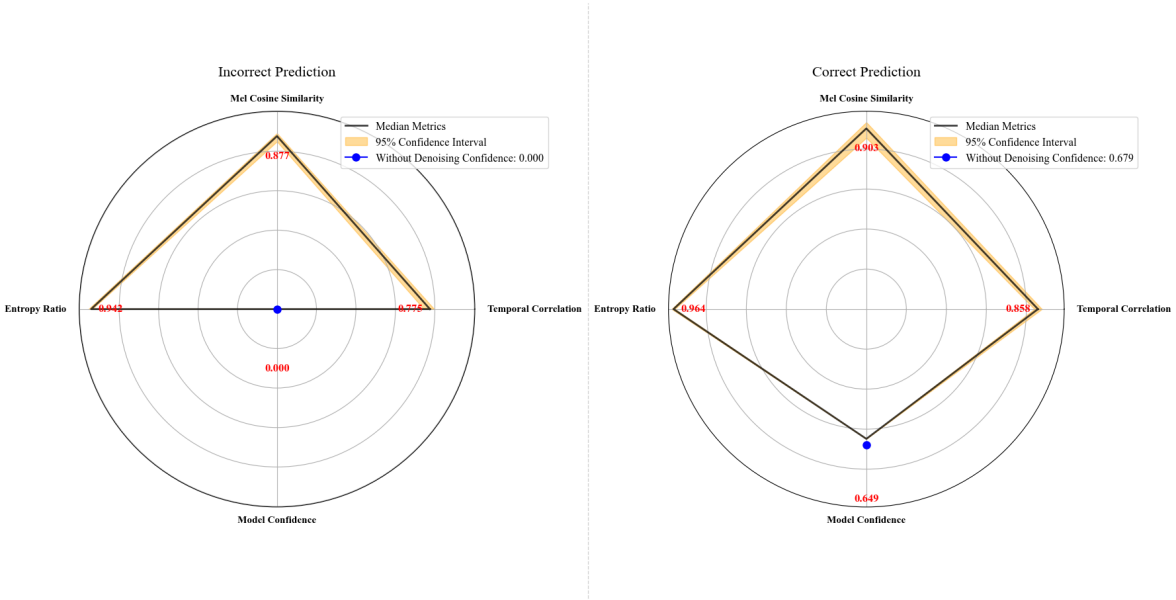


FIG. 12: *Atalotriccus pilaris* - Comparación de métricas para predicciones correctas e incorrectas, incluyendo intervalos de confianza del 95 %.

Los resultados de la prueba de Mann-Whitney U para los niveles de confianza fueron los siguientes:

- **Estadístico U**: 2621
- **Valor p**: 0.961

Dado que el valor p ($p = 0{,}961$) es considerablemente mayor que el nivel de significancia de 0.05, no se encontró evidencia estadísticamente significativa para afirmar que la confianza del modelo en los audios *denoised* es mayor que en los audios originales en el caso de predicciones correctas. Esto indica que no hay diferencias significativas en los niveles de confianza entre los dos grupos, sugiriendo que el proceso de *denoising* no afectó significativamente la confianza del modelo en este caso.

VI-B2) Gráfico de Barras de Proporciones: Para la especie *Atalotriccus pilaris*, el gráfico de proporciones revela que la proporción de clasificaciones correctas disminuye al aplicar denoising, bajando de un $29{,}6\,\%$ en los audios originales a un $13{,}3\,\%$ en los audios procesados. Este resultado sugiere que, aunque el denoising ayuda a reducir el ruido en la señal, también podría estar eliminando detalles acústicos importantes para la identificación precisa de esta especie. Sin embargo, en el análisis previo de los casos correctamente clasificados, no se observa un aumento significativo en la confianza del modelo tras el denoising. Esto indica que, aunque el proceso de denoising simplifica la señal, no necesariamente mejora la certeza del modelo en las predicciones correctas para esta especie.

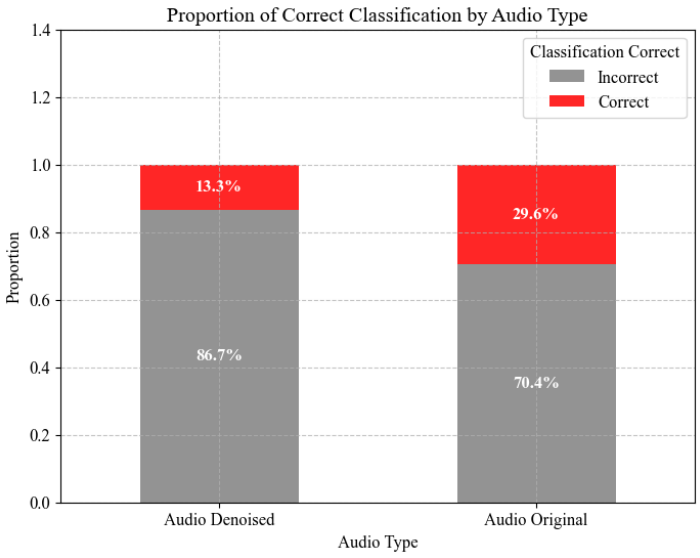


FIG. 13: Proporción de clasificaciones correctas e incorrectas para *Atalotriccus pilaris* en los audios originales y procesados con denoising (402 Muestras).

VI-B3) Resultados para Pitangus sulphuratus: En el caso de las predicciones incorrectas de *Pitangus sulphuratus*, la correlación temporal alcanza un valor mediano de $0{,}756$, lo que sugiere una pérdida moderada de la estructura temporal tras el denoising. La similitud coseno en el dominio Mel tiene un valor de $0{,}860$, lo que indica una conservación limitada de la estructura frecuencial. La razón de entropía es de $0{,}939$, lo cual refleja una reducción en la complejidad, siendo tal vez excesiva para la identificación de esta especie. La confianza del modelo en estos casos es de $0{,}0$ ya que estos son casos en donde no se detecto la especie correcta.

Para las predicciones correctas, la correlación temporal mejora, alcanzando un valor mediano de $0{,}794$, lo que implica una preservación mas robusta de la estructura temporal de las vocalizaciones de *Pitangus sulphuratus*. La similitud coseno en el dominio Mel incrementa a $0{,}909$, destacando que retiene las características frecuenciales esenciales después del denoising. La razón de entropía aumenta levemente a $0{,}942$, lo cual ayuda a reducir el ruido manteniendo un nivel de complejidad adecuado. La confianza en estos casos es elevada, con una mediana de $0{,}63$, siendo esta superior a la referencia de $0{,}389$, indicando que el proceso de denoising facilita una clasificación más precisa y fiable de esta especie.

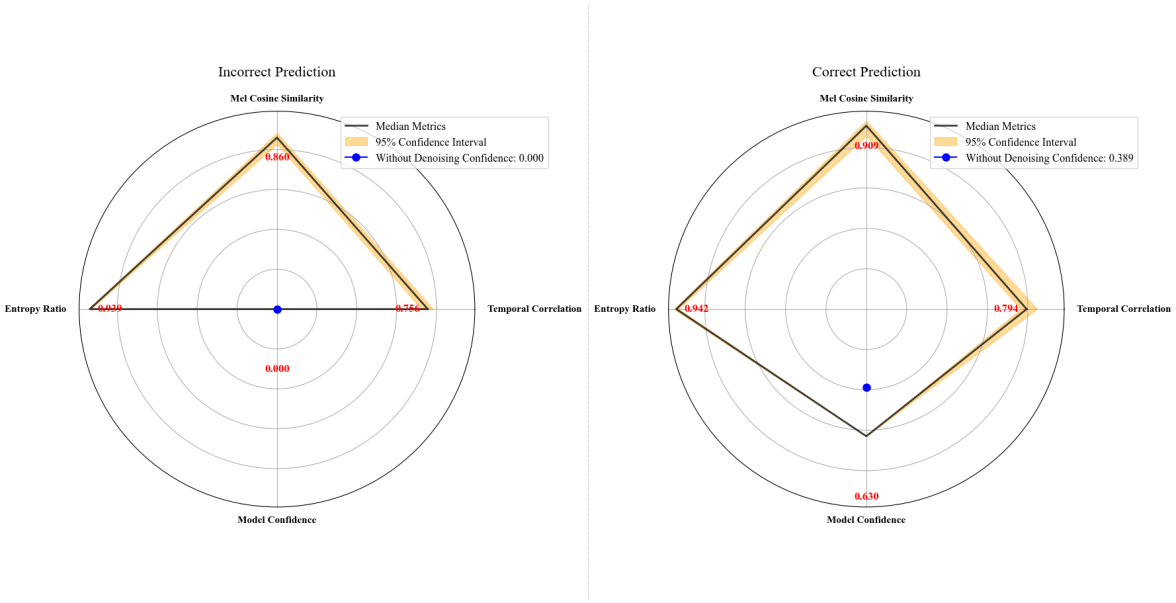


FIG. 14: *Pitangus sulphuratus* - Comparación de métricas para predicciones correctas e incorrectas, incluyendo intervalos de confianza del 95 %.

Los resultados de la prueba de Mann-Whitney U para los niveles de confianza fueron los siguientes:

- **Estadístico U**: 9400.0
- **Valor p**: 0.0019

Dado que el valor p ($p = 0{,}0019$) es menor que el nivel de significancia de 0.05, se encontró evidencia estadísticamente significativa para afirmar que la confianza del modelo en los audios *denoised* es mayor que en los audios originales en el caso de predicciones correctas. Este resultado sugiere que el proceso de *denoising* aumenta de forma significativa la confianza del modelo en sus predicciones correctas.

VI-B4) Gráfico de Barras de Proporciones: En el caso de *Pitangus sulphuratus*, la proporción de clasificaciones correctas también disminuye al aplicar denoising, pasando de un $33{,}9\,\%$ en los audios originales a un $23{,}7\,\%$ en los audios denoised. Este descenso en la proporción de aciertos sugiere que la técnica de reducción de ruido podría estar suprimiendo ciertas características frecuenciales o temporales necesarias para identificar esta especie con precisión. Sin embargo, el análisis de confianza en los casos correctamente clasificados indica que el modelo muestra una mayor seguridad en estas predicciones, lo que sugiere que el denoising mejora la claridad de la señal en aquellos casos en los que el modelo logra identificar correctamente a la especie.

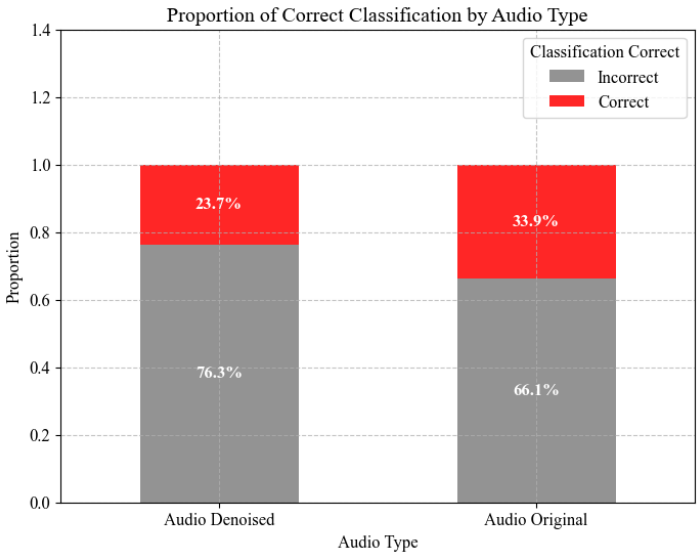


FIG. 15: Proporción de clasificaciones correctas e incorrectas para *Pitangus sulphuratus* en los audios originales y procesados con denoising (440 Muestras).

VI-B5) Resultados para Tolmomyias flaviventris: Para *Tolmomyias flaviventris*, las predicciones incorrectas presentan una correlación temporal con un valor mediano de $0{,}779$, sugiriendo una pérdida considerable de la estructura temporal en la señal denoised. La similitud coseno en el dominio Mel es de $0{,}878$, mostrando una conservación limitada de las características frecuenciales necesarias para la identificación. La razón de entropía es de $0{,}941$, indicando que la técnica de denoising reduce la complejidad, pero posiblemente elimina también detalles importantes para la clasificación. La confianza del modelo es de $0{,}0$ ya que estos son casos en donde no se detecto la especie correcta.

En los casos de predicción correcta, la correlación temporal mejora significativamente, con un valor mediano de $0{,}857$, lo que demuestra una buena preservación de la estructura temporal en las vocalizaciones de *Tolmomyias flaviventris*. La similitud coseno en el dominio Mel aumenta a $0{,}913$, lo que implica que la estructura frecuencial es mantenida de manera efectiva por el denoising en estos casos. La razón de entropía sube a $0{,}96$, manteniendo un balance adecuado entre reducción de ruido y complejidad de la señal. La confianza en la clasificación correcta es mayor, alcanzando una mediana de $0{,}524$ superando la referencia de $0{,}421$, lo que sugiere que el denoising permite una identificación precisa de la especie en condiciones de ruido reducido.

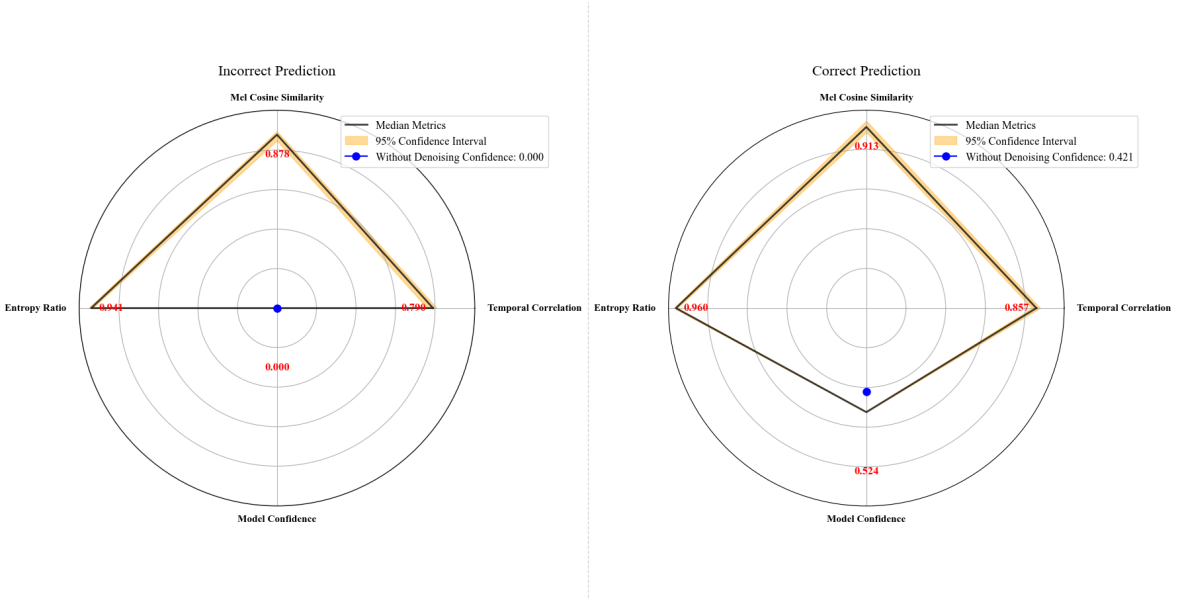


FIG. 16: *Tolmomyias flaviventris* - Comparación de métricas para predicciones correctas e incorrectas, incluyendo intervalos de confianza del 95 %.

Los resultados de la prueba de Mann-Whitney U para los niveles de confianza fueron los siguientes:

- **Estadístico U**: 2999
- **Valor p**: 0.0235

Dado que el valor p ($p = 0{,}0235$) es menor que el nivel de significancia de 0.05, se encontró evidencia estadísticamente significativa para afirmar que la confianza del modelo en los audios *denoised* es mayor que en los audios originales en el caso de predicciones correctas. Esto sugiere que el proceso de *denoising* contribuye significativamente a aumentar la confianza del modelo en sus predicciones.

VI-B6) Gráfico de Barras de Proporciones: Para *Tolmomyias flaviventris*, se observa un patrón similar: la proporción de clasificaciones correctas disminuye de $22{,}9\,\%$ en los audios originales a $14{,}9\,\%$ en los audios denoised. Esto sugiere que, aunque el denoising facilita la reducción de ruido, también puede estar suprimiendo detalles importantes para la correcta identificación de esta especie. No obstante, en los casos de predicción correcta, la confianza del modelo es mayor en los audios procesados, lo que sugiere que, cuando el modelo logra clasificar correctamente, el denoising aporta una mayor claridad y certeza en la predicción.

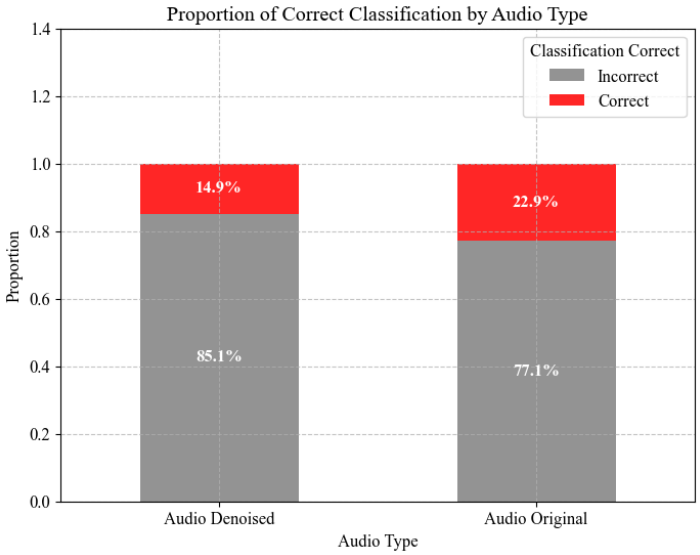
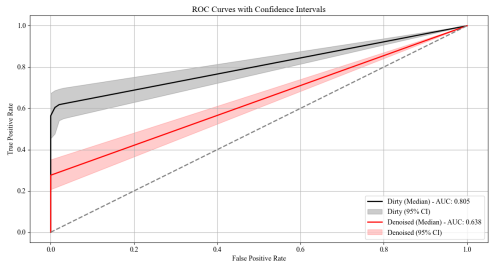


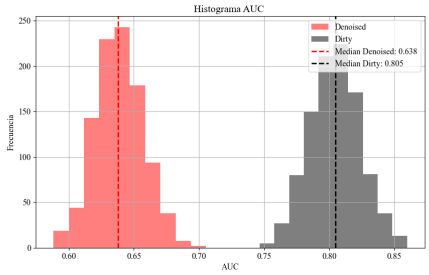
FIG. 17: Proporción de clasificaciones correctas e incorrectas para *Tolmomyias flaviventris* en los audios originales y procesados con denoising (384 Muestras).

VI-B7) Análisis de la Curva ROC: En esta sección, se presentan los resultados de las curvas ROC y los histogramas de AUC para tres especies: *Atalotriccus pilaris*, *Pitangus sulphuratus* y *Tolmomyias flaviventris*. Las curvas ROC muestran el rendimiento del clasificador en condiciones de señal original (Treshold=0) y señal denoised (Threshold=1), evaluando la capacidad del modelo para distinguir entre señales relevantes y ruido. En general, se observa una disminución en el AUC en las curvas ROC para las señales denoised en comparación con las originales, lo cual indica una reducción en la sensibilidad del clasificador tras aplicar el proceso de denoising. Esta tendencia también se evidencia

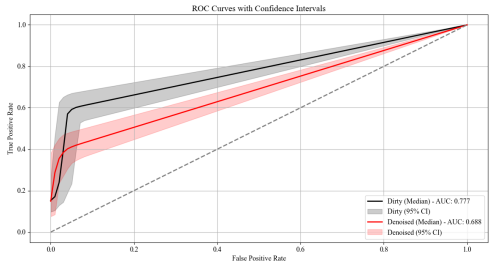
en los histogramas de AUC, donde las distribuciones para las señales denoised presentan medianas más bajas y una mayor dispersión en comparación con las señales originales. Este comportamiento sugiere que, aunque el proceso de denoising mejora la claridad de la señal, podría estar eliminando componentes acústicos importantes para la correcta clasificación, afectando negativamente el rendimiento del modelo en la detección de las especies objetivo. Estos resultados resaltan la importancia de un equilibrio cuidadoso entre la eliminación de ruido y la preservación de características clave de la señal, para evitar una degradación en la precisión de los modelos bioacústicos.



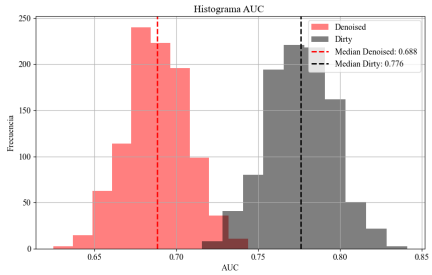
(a) Curva ROC para Atalotriccus pilaris



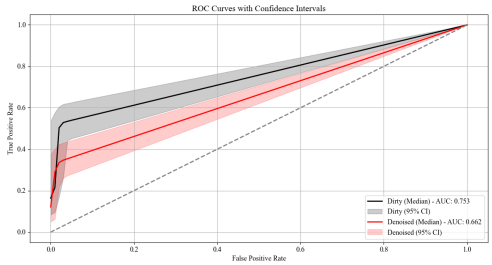
(b) Histograma AUC para Atalotriccus pilaris



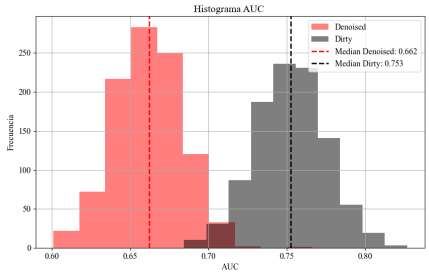
(c) Curva ROC para Pitangus sulphuratus



(d) Histograma AUC para Pitangus sulphuratus



(e) Curva ROC para Tolmomyias flaviventris



(f) Histograma AUC para Tolmomyias flaviventris

FIG. 18: Comparación de curvas ROC y distribuciones AUC para diferentes especies: Atalotriccus pilaris, Pitangus sulphuratus y Tolmomyias flaviventris. Cada pareja muestra la curva ROC junto con el histograma de AUC para cada especie, comparando datos originales y denoised.

ANUROS

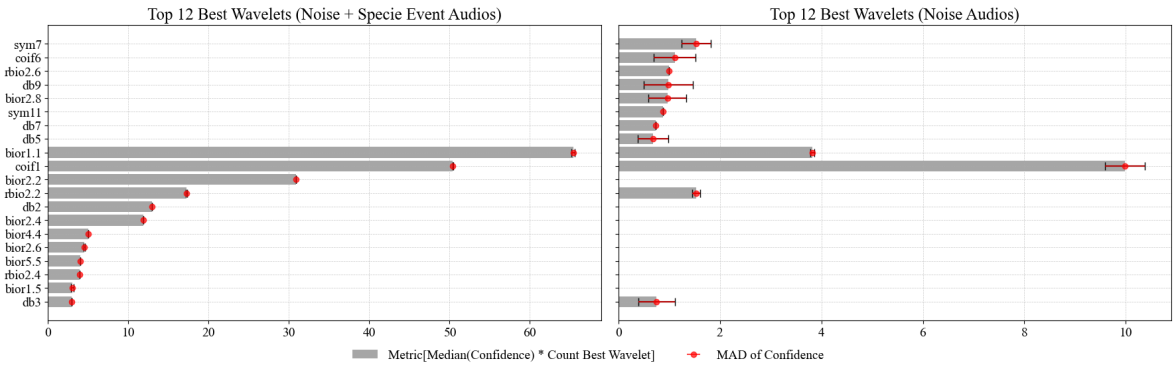


FIG. 19: Top 12 wavelets seleccionadas en los experimentos con anuros para diferentes condiciones de ruido.

El gráfico de la Figura 19 presenta las *top* 12 wavelets seleccionadas en dos escenarios experimentales utilizando la ecuación 13: *Noise + Specie Event Audios*, correspondiente a señales con ruido ambiental y eventos de llamadas de anuros, y *Noise Audios*, que representa señales de solo ruido ambiental. En el caso de *Noise + Specie Event Audios*, las wavelets destacadas muestran una alta capacidad para preservar características bioacústicas críticas, esenciales para la identificación y clasificación de especies en entornos acústicamente complejos. Por el contrario, las wavelets seleccionadas en *Noise Audios* optimizan la atenuación de ruido en ausencia de eventos de interés, lo cual es efectivo para una eliminación general de ruido sin priorizar la conservación de detalles específicos de las llamadas de anuros.

Existen varias estrategias para seleccionar el conjunto óptimo de wavelets: la unión de ambas selecciones incluiría wavelets con efectividad en al menos uno de los contextos, proporcionando versatilidad, aunque sin garantizar la preservación de los eventos bioacústicos. La selección de wavelets exclusivas del conjunto *Noise Audios* priorizaría la reducción de ruido en general, lo cual podría conllevar la pérdida de información relevante en presencia de señales bioacústicas. Por otro lado, centrarse en las wavelets exclusivas de *Noise + Specie Event Audios* como *bior2.2, db2, bior2.4, bior4.4, bior2.6, bior5.5, rbio2.4 y bior1.5* permite enfocarse en aquellas transformadas que son especialmente eficaces en la separación de ruido mientras preservan las características acústicas de interés en el contexto de eventos de anuros.

Optar por las wavelets exclusivas de *Noise + Specie Event Audios* constituye una estrategia óptima para maximizar la conservación de detalles bioacústicos específicos al tiempo que se reduce el ruido ambiental. Estas wavelets demuestran una especialización en la preservación de señales relevantes, manteniendo la integridad de los eventos de interés en entornos ruidosos. Esta elección es clave en estudios bioacústicos de anuros, donde se requiere

un equilibrio entre reducción de ruido y preservación de las características de la señal, mejorando así la precisión y fiabilidad en la identificación y análisis de especies en contextos de monitoreo ambiental.

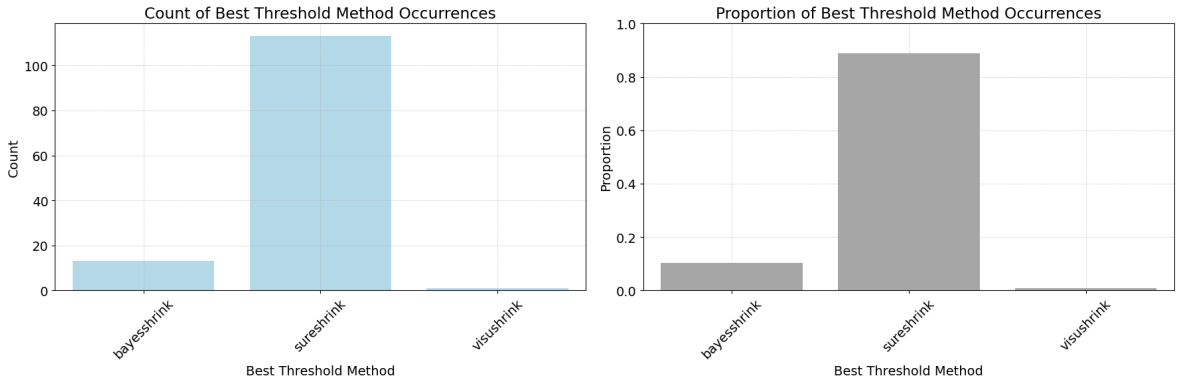


FIG. 20: Frecuencia y proporción de selección de cada método de umbralización (bayesshrink, sureshrink, visushrink) en el proceso de eliminación de ruido en señales de anuros.

El gráfico de la Figura 20 muestra la frecuencia de selección de los métodos de umbralización *bayesshrink*, *sureshrink* y *visushrink* en el proceso de eliminación de ruido aplicado a datos de señales bioacústicas de anuros. En el panel izquierdo, se observa que el método *sureshrink* es claramente dominante, con un número de ocurrencias muy superior en comparación con los otros métodos, cuya frecuencia es marginal. Esta tendencia se confirma en el panel derecho, donde la proporción de selección de *sureshrink* se aproxima al 100 %, indicando que prácticamente todos los casos donde la predicción fue correcta emplearon este método de umbralización.

Estos resultados sugieren que *sureshrink* es el método más eficaz en la configuración experimental utilizada para señales de anuros, ya que se destaca en su capacidad para preservar características bioacústicas importantes al mismo tiempo que reduce el ruido. Para optimizar el costo computacional en análisis futuros, se decide emplear las wavelets seleccionadas previamente en el conjunto *Noise + Specie Event Audios* (*bior2.2*, *db2*, *bior2.4*, *bior4.4*, *bior2.6*, *bior5.5*, *rbio2.4* y *bior1.5*) en combinación con el método de umbralización *sureshrink*. Esto garantiza la eficiencia y la conservación de los detalles acústicos necesarios para la identificación y clasificación de especies de anuros en entornos ruidosos, mejorando así la precisión en estudios de monitoreo ambiental de especies.

VI-C. Ruido vs. Vocalizaciones de Especies

Este estudio examina el caso de Ruido vs. Vocalizaciones de Especies en anuros, evaluando la capacidad del modelo de clasificación bioacústica para distinguir entre segmentos de audio que contienen vocalizaciones de anuros y aquellos que solo contienen ruido ambiental. Se emplearon datos del conjunto AnuraSet, que ya incluyen etiquetas de ruido y vocalizaciones, asegurando una clasificación homogénea y precisa entre ruido y vocalizaciones.

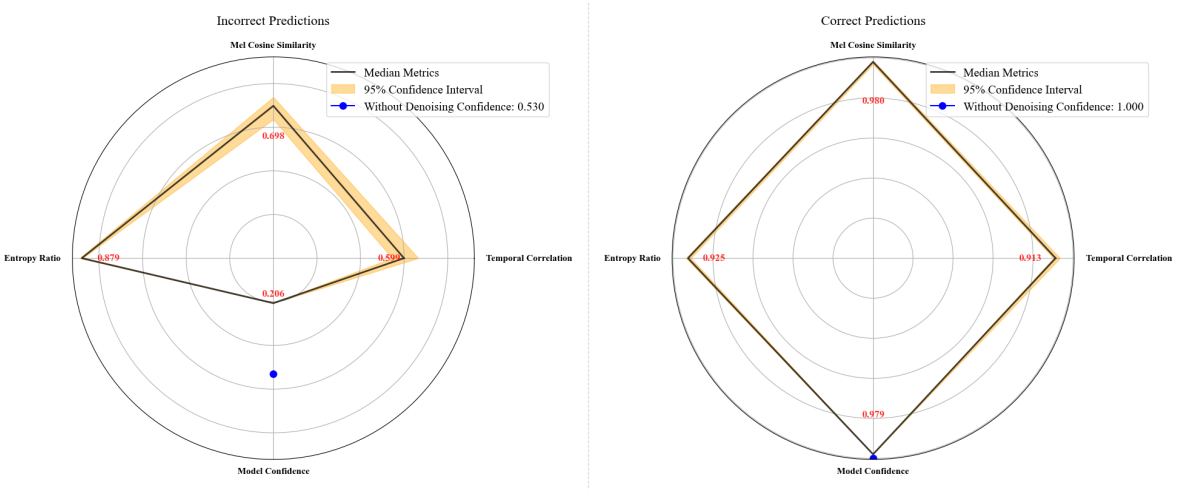


FIG. 21: Gráfico radar de las métricas de evaluación para predicciones correctas e incorrectas en el caso de los anuros, mostrando la similitud coseno en el dominio Mel, la correlación temporal, la razón de entropía y la confianza del modelo.

VI-C1) Gráficos Radar: Evaluación de Métricas de Denoising: En el gráfico de radar (Figura 21) correspondiente a las predicciones incorrectas, observamos que la **similitud coseno en el dominio Mel** tiene un valor mediano de aproximadamente 0.698, mientras que la **correlación temporal** es de 0.599. Estos valores sugieren que, en los casos de predicción incorrecta, el proceso de *denoising* no logra preservar adecuadamente la estructura temporal y frecuencial de las vocalizaciones de anuros, lo que implica que el ruido ambiental interfiere con las características relevantes para la clasificación. Además, se puede observar que la **confianza del modelo** disminuye drásticamente después del procesamiento de eliminación de ruido, pasando de un valor de referencia sin denoising de 0.530 (indicando una alta seguridad en predicciones incorrectas) a un valor mediano de 0.206. Este fenómeno sugiere que el modelo, al procesar las señales denoised, reduce significativamente su nivel de confianza en casos donde anteriormente estaba erróneamente seguro de sus predicciones. La **razón de entropía** es de 0.879, indicando una reducción moderada de la información en la señal denoised en comparación con la señal original, lo cual muestra una simplificación parcial de la señal. En resumen, el denoising no solo afecta la estructura frecuencial y temporal de las señales, sino que también induce un ajuste en la seguridad del modelo, reduciendo la confianza en las predicciones que anteriormente eran incorrectas.

En contraste, el gráfico radar para las predicciones correctas muestra una mejora significativa en las métricas clave. La **similitud coseno en el dominio Mel** alcanza un valor mediano de 0.980 y la **correlación temporal** sube a 0.913, lo que indica que el modelo retiene mejor las características temporales y frecuenciales tras el *denoising*, permitiendo una identificación más precisa de las vocalizaciones de anuros. La **confianza del modelo** es también considerablemente mayor en estos casos, con un valor mediano de 0.979, lo cual sugiere que el proceso de reducción de ruido facilita una clasificación más confiable. La

razón de entropía es de 0.925, lo que indica que el *denoising* logra reducir la complejidad de la señal de ruido manteniendo la integridad de las vocalizaciones de los anuros.

Se realizó una prueba de Mann-Whitney U para comparar la confianza del modelo entre los audios *denoised* y los audios originales en predicciones correctas. Los resultados de la prueba fueron los siguientes:

- **Estadístico U**: 5052.5
- **Valor p**: 0.9999

Dado que el valor p ($p = 0{,}9999$) es significativamente mayor que el nivel de significancia de 0.05, no se encontró evidencia estadísticamente significativa para afirmar que la confianza del modelo en los audios *denoised* es mayor que en los audios originales en el caso de predicciones correctas. Este resultado indica que, aunque se observa un incremento en la confianza en los audios denoised, este aumento no es estadísticamente significativo en comparación con los audios originales.

VI-C2) Gráfico de Barras de Proporciones: El gráfico de barras de proporciones (Figura 22) muestra que el "Audio Denoised" presenta un 21.8 % de predicciones correctas y un 78.2 % de predicciones incorrectas, mientras que el "Audio Original" alcanza un 29.1 % de predicciones correctas y un 70.9 % de predicciones incorrectas. En este caso, el proceso de *denoising* no mejora las métricas de confianza, ademas de que la proporción de predicciones correctas es menor en el audio denoised en comparación con el audio original. Esto sugiere que, aunque la reducción de ruido ayuda a disminuir la interferencia ambiental, también puede eliminar algunas características relevantes de las vocalizaciones, afectando la precisión del modelo en la clasificación de las señales de anuros.

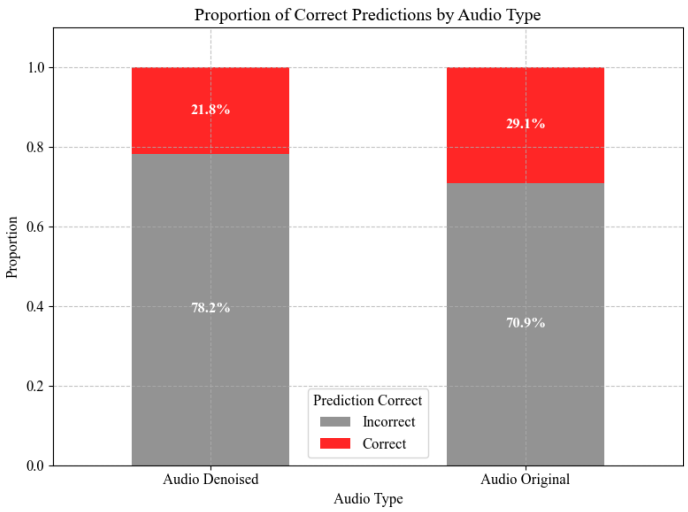


FIG. 22: Proporción de predicciones correctas e incorrectas por tipo de audio (Audio Denoised y Audio Original) en el caso de los anuros (523 Muestras).

VI-C3) Análisis de la Distancia de Jaccard: La distancia de Jaccard se calculó para comparar la precisión en la clasificación multietiqueta entre audios con y sin eliminación de ruido en el caso de los anuros. Los resultados indican que la distancia de Jaccard para el caso con eliminación de ruido es de 0.5455, mientras que para el caso sin eliminación de ruido es de 0.4118. Este incremento en la distancia de Jaccard sugiere una menor coincidencia entre las etiquetas predichas y las etiquetas reales en el caso de denoising, en comparación con el caso sin reducción de ruido.

Este resultado implica que, aunque el proceso de denoising ayuda a reducir la interferencia de ruido ambiental, también podría estar eliminando o atenuando características importantes necesarias para la identificación precisa de las vocalizaciones de anuros. En consecuencia, la eliminación de ruido podría estar afectando la precisión del modelo en la clasificación multietiqueta, disminuyendo la coincidencia entre las etiquetas predichas y las etiquetas reales. Por lo tanto, aunque el denoising contribuye a una señal más clara, su impacto en la coincidencia de etiquetas sugiere una posible limitación en la precisión de la clasificación en este contexto.

VI-C4) Análisis de la Curva ROC y Distribuciones de AUC: La curva ROC (Figura 23) compara el desempeño de clasificación en términos de tasas de verdaderos positivos frente a tasas de falsos positivos. Los resultados muestran que el audio original (Dirty) alcanza un valor de AUC mediano de 0.919, mientras que el audio denoised (Denoised)) alcanza un AUC mediano de 0.852. Esta disminución en el AUC sugiere que el proceso de denoising, aunque reduce el ruido de fondo, podría estar afectando la capacidad del modelo para detectar correctamente las características de las señales bioacústicas relevantes.

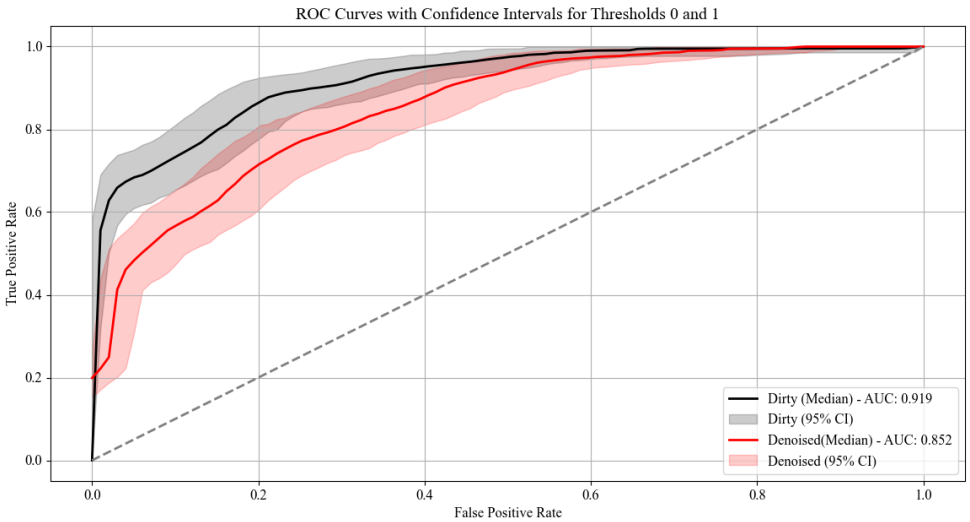


FIG. 23: Curva ROC con intervalos de confianza para los casos Dirty (Señal Origianal) y Denoised (Señal Procesada). El AUC mediano es mayor para el audio original (0.919) en comparación con el audio denoised (0.852), lo que sugiere un impacto negativo del proceso de denoising en la precisión de clasificación.

Además, se realizaron análisis de las distribuciones de AUC generadas mediante el método bootstrap (Figura 24). Los histogramas muestran una distribución de los AUC más alta y con menor varianza para el audio original, en comparación con el audio denoised. La mediana del AUC para el audio original se encuentra en 0.919, mientras que para el audio denoised está en 0.850. Esto refuerza la hipótesis de que la eliminación de ruido, aunque mejora la claridad de la señal, puede estar atenuando características importantes para la clasificación precisa, lo que se refleja en un AUC menor y una menor confianza en la clasificación. Ademas se realizó una prueba de Wilcoxon para comparar los valores de AUC de las señales con y sin denoising, obteniéndose un estadístico de prueba de 0.0 y un valor p de 1.0. Dado que el valor p es significativamente mayor que el umbral común de significancia (0.05), no se puede rechazar la hipótesis nula. Esto indica que no hay evidencia suficiente para concluir que los AUC del audio denoised sean mayores que los del audio original. En otras palabras, el proceso de denoising no proporciona una mejora estadísticamente significativa en el AUC en comparación con el audio original.

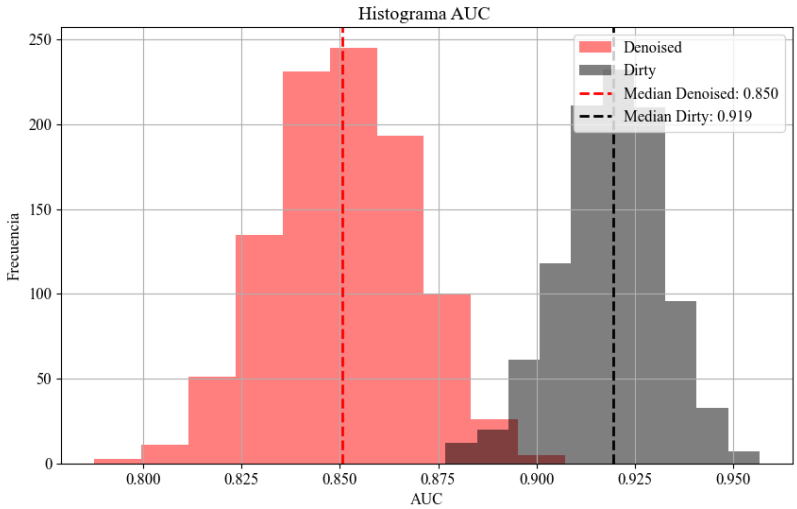


FIG. 24: Distribuciones de los valores de AUC obtenidos mediante bootstrap para los casos con y sin eliminación de ruido. Se observa que el AUC mediano es mayor para el audio original (0.919) en comparación con el audio denoised (0.850), lo que indica una mayor precisión de clasificación sin el proceso de reducción de ruido.

VI-D. Vocalizaciones de Especies Frecuentes vs. Ruido

En este escenario, el análisis se centró en evaluar la capacidad del modelo para identificar y diferenciar las vocalizaciones pertenecientes a las tres especies más abundantes en el conjunto de datos: *DENMIN*, *LEPPOD* y *PHYALB*, frente a las vocalizaciones de otras especies, que en este contexto se consideraron como ruido. con el objetivo de evaluar la habilidad del modelo para priorizar y clasificar correctamente las especies de interés en presencia de ruido. La evaluación considera tanto el rendimiento en audios originales como en aquellos a los que se les ha aplicado denoising mediante técnicas basadas en wavelets, utilizando métricas clave como la similitud coseno en el dominio Mel, la correlación temporal, la razón de entropía y la confianza del modelo en las predicciones. Además, se realizó una prueba de hipótesis para determinar si existe una diferencia significativa en la confianza del modelo entre audios denoised y originales.

VI-D1) Resultados para DENMIN: Los gráficos de radar en la Figura 25 muestran las métricas de rendimiento para las predicciones incorrectas y correctas en la especie *DENMIN*. En las predicciones incorrectas, las métricas de rendimiento son menores, con una similitud coseno en el dominio Mel de $0{,}861$ y una correlación temporal de $0{,}736$. El ratio de entropía en las predicciones incorrectas es de $0{,}867$, indicando una reducción moderada de la complejidad de la señal tras el proceso de denoising. Sin embargo, el valor relativamente bajo del ratio de entropía sugiere que el proceso de denoising podría estar eliminando información significativa, lo cual contribuye a una menor precisión en las

predicciones. Además, se observa una disminución notable en la **confianza del modelo** para las predicciones incorrectas tras el denoising, reduciéndose de un valor de referencia de $0{,}988$ (sin denoising) a $0{,}335$. Esto refleja que el modelo estaba erróneamente seguro en estos casos antes del procesamiento, y el denoising actúa como un regulador que reduce esta confianza en predicciones incorrectas. Por otro lado, en las predicciones correctas, no hay una pérdida significativa de confianza, manteniéndose en valores altos tanto antes como después del denoising (de $0{,}999$ a $0{,}977$), lo cual demuestra que el denoising preserva adecuadamente las características relevantes para clasificaciones correctas, asegurando la fiabilidad del modelo en estos casos.

Por otro lado, en las predicciones correctas, se observa una mejor preservación de las características de la señal. La similitud coseno en el dominio Mel alcanza un valor de $0{,}977$, indicando una alta retención de las características frecuenciales relevantes después del denoising. La correlación temporal también es elevada en las predicciones correctas, con un valor de $0{,}899$, lo cual sugiere que el proceso de denoising ha logrado conservar la estructura temporal de las vocalizaciones. Además, el ratio de entropía en las predicciones correctas es de $0{,}907$, lo que implica una reducción eficiente de ruido mientras se mantiene la complejidad esencial de la señal para una correcta identificación de la especie.

En cuanto a la confianza del modelo, en las predicciones correctas la confianza promedio alcanza un valor cercano a $0{,}977$, mientras que la confianza de referencia para audios sin denoising es de $0{,}999$. Esta ligera diferencia indica que, aunque el denoising introduce ciertos cambios en la señal, la confianza del modelo se mantiene alta y cercana a la confianza de referencia.

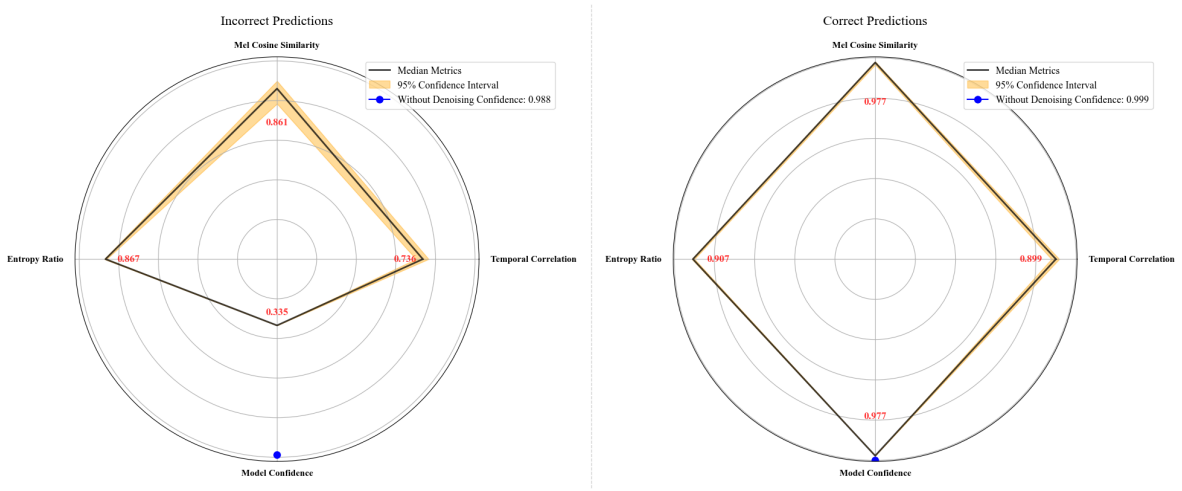


FIG. 25: Gráficos de radar de las métricas de evaluación para las predicciones incorrectas (izquierda) y correctas (derecha) en la especie *DENMIN*.

VI-D2) Gráfico de Barras de Proporciones: En cuanto a la proporción de predicciones correctas y erróneas, el análisis, representado en la Figura 26, muestra una diferencia notable entre los audios originales y los denoised. Para los audios denoised, la proporción de predicciones correctas es de $49{,}4\,\%$, mientras que para los audios originales esta proporción aumenta a $74{,}4\,\%$. Este resultado sugiere que, aunque el denoising mejora la limpieza de la señal, puede introducir modificaciones que dificulten la identificación correcta de la especie en algunos casos.

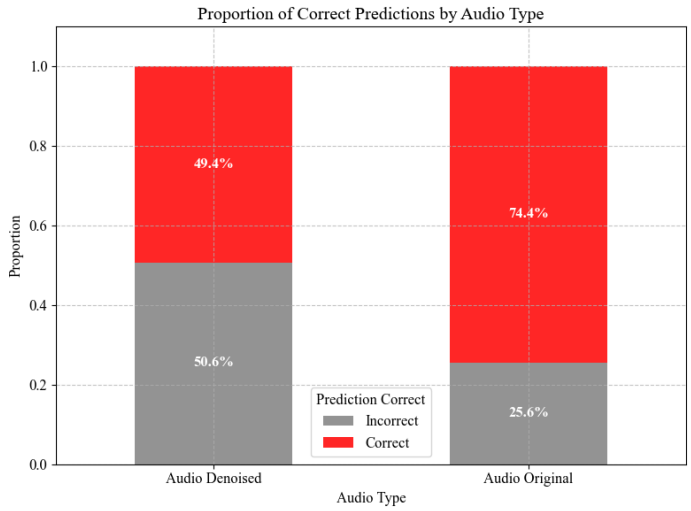


FIG. 26: Proporción de predicciones correctas y erróneas para la especie *DENMIN* en audios denoised y originales (640 Muestras).

Los resultados de la prueba de Mann-Whitney U para los niveles de confianza fueron los siguientes:

- **Estadístico U**: 5052.5
- **Valor p**: 0.999

Dado que el valor p ($p = 0{,}999$) es mayor que el nivel de significancia de 0.05, no se encontró evidencia estadísticamente significativa para afirmar que la confianza del modelo en los audios *denoised* es mayor que en los audios originales en el caso de predicciones correctas.

VI-D3) Resultados para LEPPOD: Los gráficos de radar en la Figura 27 muestran las métricas de rendimiento para las predicciones incorrectas y correctas en la especie *LEPPOD*. En las predicciones incorrectas, las métricas de rendimiento son menores, con una similitud coseno en el dominio Mel de $0{,}772$ y una correlación temporal de $0{,}706$. El ratio de entropía en las predicciones incorrectas es de $0{,}884$, indicando una reducción moderada de la complejidad de la señal tras el proceso de denoising. Sin embargo, el

valor relativamente bajo del ratio de entropía sugiere que el proceso de denoising podría estar eliminando información significativa, lo cual contribuye a una menor precisión en las predicciones. Además, se observa una disminución notable en la **confianza del modelo** para las predicciones incorrectas tras el denoising, reduciéndose de un valor de referencia sin denoising de $0{,}984$ a $0{,}353$. Esto sugiere que el modelo estaba erróneamente seguro en estas predicciones antes del procesamiento, y el denoising actúa como un regulador que disminuye dicha confianza en predicciones incorrectas. Por otro lado, en las predicciones correctas, no hay una pérdida significativa de confianza, manteniéndose en valores altos tanto antes como después del denoising (de $0{,}996$ a $0{,}971$). Esto evidencia que el proceso de denoising logra preservar adecuadamente las características relevantes para clasificaciones correctas, asegurando la fiabilidad del modelo en estos casos.

Por otro lado, en las predicciones correctas, se observa una mejor preservación de las características de la señal. La similitud coseno en el dominio Mel alcanza un valor de $0{,}971$, indicando una alta retención de las características frecuenciales relevantes después del denoising. La correlación temporal también es elevada en las predicciones correctas, con un valor de $0{,}903$, lo cual sugiere que el proceso de denoising ha logrado conservar la estructura temporal de las vocalizaciones. Además, el ratio de entropía en las predicciones correctas es de $0{,}927$, lo que implica una reducción eficiente de ruido mientras se mantiene la complejidad esencial de la señal para una correcta identificación de la especie. En cuanto a la confianza del modelo, en las predicciones correctas la confianza promedio alcanza un valor cercano a $0{,}884$, mientras que la confianza de referencia para audios sin denoising es de $0{,}996$. Esta ligera diferencia indica que, aunque el denoising introduce ciertos cambios en la señal, la confianza del modelo se mantiene alta y cercana a la confianza de referencia.

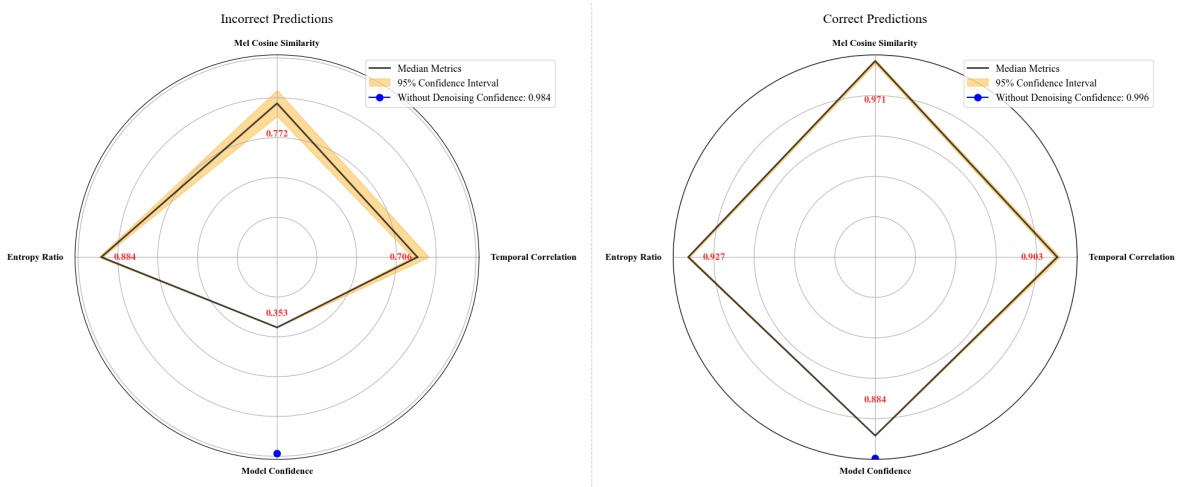


FIG. 27: Gráficos de radar de las métricas de evaluación para las predicciones incorrectas (izquierda) y correctas (derecha) en la especie *LEPPOD*.

VI-D4) Gráfico de Barras de Proporciones: En cuanto a la proporción de predicciones correctas y erróneas, el análisis, representado en la Figura 28, muestra una diferencia notable entre los audios originales y los denoised. Para los audios denoised, la proporción de predicciones correctas es de $27{,}7\,\%$, mientras que para los audios originales esta proporción aumenta a $69{,}5\,\%$. Este resultado sugiere que, aunque el denoising mejora la limpieza de la señal, puede introducir modificaciones que dificulten la identificación correcta de la especie en algunos casos.

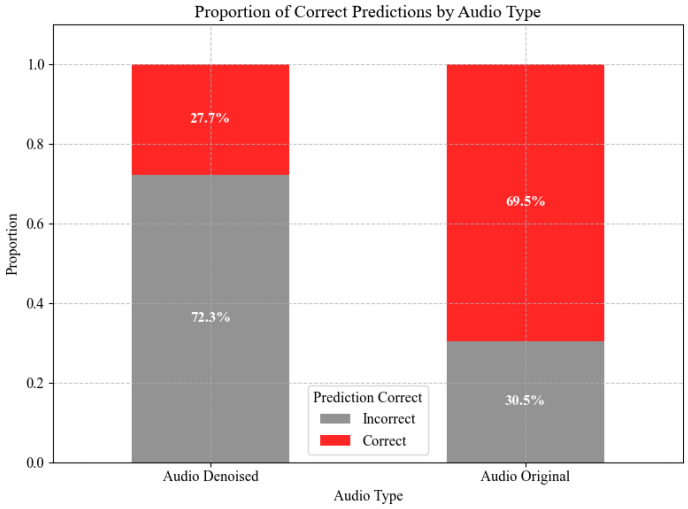


FIG. 28: Proporción de predicciones correctas y erróneas para la especie *LEPPOD* en audios denoised y originales (842 Muestras).

Los resultados de la prueba de Mann-Whitney U para los niveles de confianza fueron los siguientes:

- **Estadístico U**: 43712.5
- **Valor p**: 0.999

Dado que el valor p ($p = 0{,}999$) es mayor que el nivel de significancia de 0.05, no se encontró evidencia estadísticamente significativa para afirmar que la confianza del modelo en los audios *denoised* es mayor que en los audios originales en el caso de predicciones correctas.

VI-D5) Resultados para PHYALB: Los gráficos de radar en la Figura 29 muestran las métricas de rendimiento para las predicciones incorrectas y correctas en la especie *PHYALB*. En las predicciones incorrectas, las métricas de rendimiento son menores, con una similitud coseno en el dominio Mel de $0{,}839$ y una correlación temporal de $0{,}682$. El ratio de entropía en las predicciones incorrectas es de $0{,}861$, lo cual indica una reducción moderada de la complejidad de la señal tras el proceso de denoising. Este valor sugiere que el proceso

de denoising podría estar eliminando cierta información importante, contribuyendo a una menor precisión en las predicciones. Además, se observa una disminución significativa en la **confianza del modelo** para las predicciones incorrectas tras el denoising, reduciéndose de un valor de referencia sin denoising de $0{,}984$ a $0{,}315$. Esto implica que el modelo estaba erróneamente seguro de sus predicciones antes del procesamiento, y el denoising actúa como un regulador, reduciendo dicha confianza en casos incorrectos. Por otro lado, en las predicciones correctas, la confianza del modelo permanece alta tras el denoising, pasando de $1{,}000$ sin denoising a $0{,}990$, lo que demuestra que el proceso de denoising no afecta significativamente las características clave necesarias para la correcta clasificación. Estos resultados destacan el balance del denoising entre la reducción de ruido y la preservación de las características relevantes en las predicciones correctas.

Por otro lado, en las predicciones correctas, se observa una mejor preservación de las características de la señal. La similitud coseno en el dominio Mel alcanza un valor de $0{,}948$, lo que indica una alta retención de las características frecuenciales relevantes después del denoising. La correlación temporal también es elevada en las predicciones correctas, con un valor de $0{,}887$, lo cual sugiere que el proceso de denoising ha logrado conservar la estructura temporal de las vocalizaciones. Además, el ratio de entropía en las predicciones correctas es de $0{,}930$, lo que implica una reducción eficiente de ruido mientras se mantiene la complejidad esencial de la señal para una correcta identificación de la especie.

En cuanto a la confianza del modelo, en las predicciones correctas la confianza promedio alcanza un valor cercano a $0{,}990$, mientras que la confianza de referencia para audios sin denoising es de $1{,}000$. Esta ligera diferencia indica que, aunque el denoising introduce ciertos cambios en la señal, la confianza del modelo se mantiene alta y cercana a la confianza de referencia.

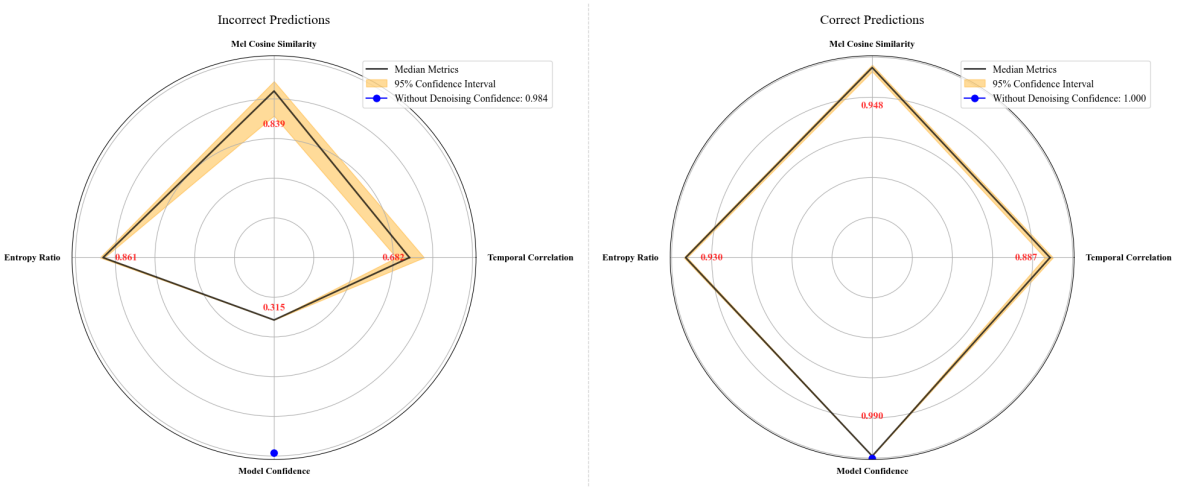


FIG. 29: Gráficos de radar de las métricas de evaluación para las predicciones incorrectas (izquierda) y correctas (derecha) en la especie *PHYALB*.

VI-D6) Gráfico de Barras de Proporciones: En cuanto a la proporción de predicciones correctas y erróneas, el análisis, representado en la Figura 30, muestra una diferencia notable entre los audios originales y los denoised. Para los audios denoised, la proporción de predicciones correctas es de $61{,}8\,\%$, mientras que para los audios originales esta proporción aumenta a $72{,}6\,\%$. Este resultado sugiere que, aunque el denoising mejora la limpieza de la señal, puede introducir modificaciones que dificulten la identificación correcta de la especie en algunos casos.

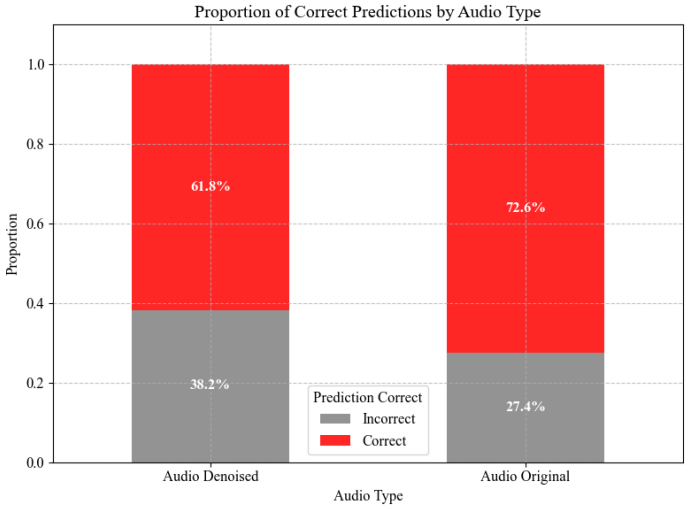


FIG. 30: Proporción de predicciones correctas y erróneas para la especie *PHYALB* en audios denoised y originales (434 Muestras).

Los resultados de la prueba de Mann-Whitney U para los niveles de confianza fueron los siguientes:

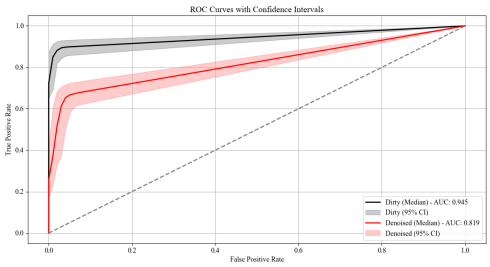
- **Estadístico U**: 26007.5
- **Valor p**: 1.0

Dado que el valor p ($p = 1{,}0$) es mayor que el nivel de significancia de 0.05, no se encontró evidencia estadísticamente significativa para afirmar que la confianza del modelo en los audios *denoised* es mayor que en los audios originales en el caso de predicciones correctas. Esto sugiere que, aunque la precisión general en las predicciones varía entre audios originales y denoised, la confianza del modelo en las predicciones correctas no cambia de forma significativa.

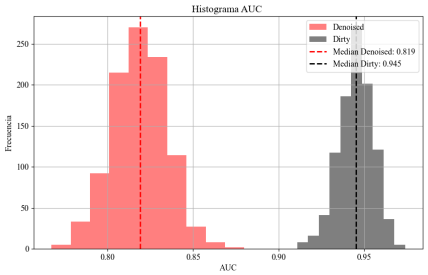
VI-D7) Análisis de la Curva ROC: La Figura 31 presenta un análisis comparativo del desempeño del modelo en la clasificación de vocalizaciones de tres especies de anuros: *DENMIN*, *LEPPOD* y *PHYALB*. Para cada especie, se muestran dos tipos de gráficos: una curva ROC y un histograma de AUC. Las curvas ROC permiten visualizar la capacidad

del modelo para diferenciar entre clases positivas y negativas en distintas configuraciones, con intervalos de confianza al 95 % para audios originales y procesados (denoised). En general, una mayor AUC indica un mejor desempeño del modelo, y la comparación entre las curvas ROC de los datos originales y denoised ayuda a identificar el impacto del proceso de denoising en la precisión de la clasificación.

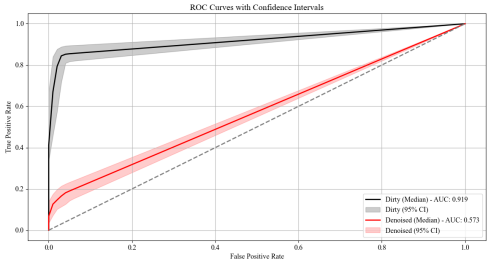
Los histogramas de AUC complementan este análisis mostrando la distribución de las áreas bajo la curva para ambas condiciones (audios originales y denoised), permitiendo observar cambios en las medianas que sugieren posibles pérdidas de información acústica relevante después del denoising. Una disminución en la mediana de AUC en los audios denoised podría indicar que el proceso de limpieza de ruido afecta la capacidad del modelo para identificar correctamente las vocalizaciones, probablemente debido a la eliminación de características acústicas críticas para la clasificación. Este análisis conjunto proporciona una visión detallada del impacto del denoising en la precisión del modelo para cada especie de anuro, destacando tanto los beneficios como las limitaciones de este proceso de limpieza en el contexto de la bioacústica.



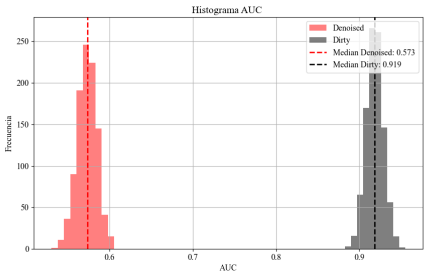
(a) Curva ROC para DENMIN



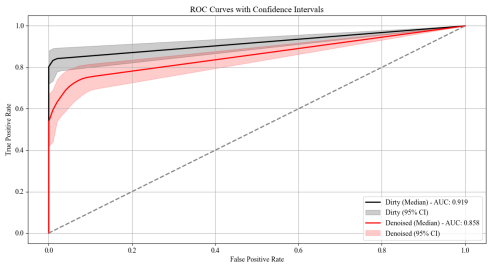
(b) Histograma AUC para DENMIN



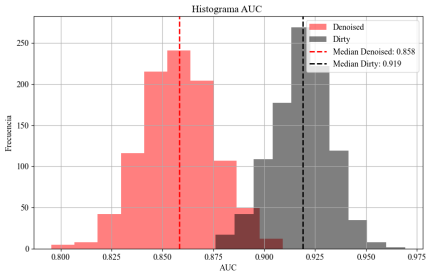
(c) Curva ROC para LEPPOD



(d) Histograma AUC para LEPPOD



(e) Curva ROC para PHYALB



(f) Histograma AUC para PHYALB

FIG. 31: Comparación de curvas ROC y distribuciones AUC para diferentes especies de anuros: DENMIN, LEPPOD y PHYALB. Cada pareja muestra la curva ROC junto con el histograma de AUC para cada especie, comparando datos originales y denoised.

VII. Conclusiones

- **Efectividad de las técnicas de reducción de ruido mediante wavelets**: La aplicación de técnicas de reducción de ruido basadas en wavelets ha mostrado ser efectiva en la eliminación de ruido ambiental en señales bioacústicas, lo cual contribuye a mejorar la claridad de las vocalizaciones de especies en condiciones acústicas complejas. Sin embargo, este proceso puede eliminar detalles acústicos esenciales para la clasificación precisa en algunos casos, indicando la necesidad de un balance entre la reducción de ruido y la preservación de información relevante, así como de ajustar los parámetros de umbralización para cada aplicación específica.

- **Impacto en la precisión y confianza del modelo**: El modelo de clasificación BirdNET experimento un incremento en la confianza de predicción en condiciones de ruido reducido, especialmente en señales donde las vocalizaciones son dominantes. Aunque el uso de técnicas como el método de umbralización *sureshrink* y wavelets seleccionadas optimizó la preservación de las características bioacústicas, en algunos casos la precisión de clasificación en las señales denoised no fue significativamente mayor que en las señales originales, subrayando la necesidad de ajustar los parámetros de denoising de acuerdo con el tipo de especie y el objetivo del monitoreo.

- Los resultados obtenidos con BirdNET destacan que la eliminación de ruido es efectiva para mejorar la discriminación entre vocalizaciones de aves y ruido ambiental, aumentando el AUC y mostrando un desempeño más confiable en entornos acústicamente complejos. Esta mejora en la capacidad de discriminación y la menor tasa de falsas alarmas sugieren que el denoising es especialmente valioso en escenarios ruidosos, como áreas urbanas o selvas, donde permite un monitoreo más eficiente y robusto a largo plazo. Aunque la técnica reduce la tasa de aciertos al suprimir algunas características sutiles de las vocalizaciones, los gráficos radar y el análisis estadístico indican que con ajustes en el proceso de denoising es posible optimizar el balance entre reducción de ruido y preservación de características bioacústicas esenciales. En consecuencia, esta técnica es viable para aplicaciones en entornos de alta interferencia, pero podría no ser adecuada en estudios que requieren captar todas las vocalizaciones posibles, como en el caso de especies en peligro o con baja densidad poblacional.

- En el modelo Anuraset, se observó que el denoising redujo drásticamente la confianza en las predicciones incorrectas, actuando como un regulador para evitar clasificaciones erróneas con alta seguridad. Además, la confianza en las predicciones correctas se mantuvo alta tras el proceso de denoising, lo que demuestra que las características relevantes para la clasificación se preservan adecuadamente en estos casos. Este equi-

librio evidencia la utilidad del denoising en escenarios con ruido significativo, donde permite aumentar la precisión general del modelo y reducir las falsas alarmas sin afectar la confiabilidad en predicciones correctas.

- **Relevancia para el monitoreo ambiental y la conservación de la biodiversidad**: La implementación de modelos de clasificación bioacústica en conjunto con métodos avanzados de procesamiento de señales, como el denoising basado en wavelets, representa un avance significativo para el monitoreo pasivo de la biodiversidad. Este enfoque permite una mayor robustez en la identificación de especies en entornos ruidosos, lo cual fortalece los esfuerzos de monitoreo y conservación de ecosistemas. Los hallazgos resaltan el potencial de estas tecnologías para mejorar el monitoreo a gran escala de la biodiversidad, proporcionando datos de alta calidad para la toma de decisiones en conservación. Futuros estudios pueden enfocar esfuerzos en el ajuste fino de los parámetros de denoising para maximizar el rendimiento en diferentes contextos de conservación.

VIII. Plan de Gestión de Datos

En el marco de este proyecto de grado, se han empleado dos fuentes de información primordiales: AnuraSet y Parex 2023. Su contribución ha sido fundamental para el avance de esta investigación al proporcionar datos de alta calidad y relevancia. Agradecemos al Instituto Humboldt por su invaluable apoyo en la facilitación del acceso y la utilización de estas fuentes, cuya intervención ha sido esencial para nuestro desarrollo y comprensión de la relevancia biológica.

Hemos adoptado una postura proactiva hacia los datos abiertos, reconociendo su capacidad para fomentar la transparencia y facilitar la difusión del conocimiento. Este enfoque no solo ha enriquecido nuestras investigaciones con información accesible y gratuita, sino que también promueve el acceso universal a datos para futuras investigaciones.

IX. Aspectos Éticos

En el contexto de este proyecto de investigación, los datos bioacústicos utilizados, provenientes de AnuraSet y Parex 2023, se consideran de interés general y acceso público, lo que exime al estudio de consideraciones éticas específicas en relación con su recolección o uso. Se ha adoptado una postura proactiva hacia el uso de datos abiertos, reconocido por su potencial para promover la transparencia y la difusión del conocimiento científico. Es crucial subrayar que cualquier mención a las fuentes de datos se realiza con el propósito de reconocer su contribución al estudio, los resultados de este son un esfuerzo investigativo y no una posición oficial del instituto.

X. Trabajos Futuros

Como trabajo futuro, se propone explorar el uso de técnicas de reducción de ruido como una estrategia de aumentación de datos para la clasificación de señales bioacústicas en proyectos similares. Este enfoque tiene el potencial de mitigar la contaminación acústica que afecta la calidad de las grabaciones, especialmente en entornos ruidosos. Al mejorar la claridad de las señales de origen mediante estas técnicas, se espera reducir el impacto del ruido ambiental y optimizar la precisión de los modelos de clasificación.

Además, a partir de la señal limpia generada mediante reducción de ruido, se podrían crear versiones contaminadas con distintos tipos de ruido. Este procedimiento permitiría entrenar modelos más robustos y adaptativos ante diversas condiciones acústicas. La creación de estas variaciones aumentaría la capacidad del modelo para generalizar en contextos ruidosos, elevando así su aplicabilidad en entornos de monitoreo real.

Integrar la reducción de ruido como parte del preprocesamiento también facilita la extracción de características relevantes de la señal bioacústica. Esto abre la puerta a experimentar con modelos avanzados, como BirdNET y AnuraSet, los cuales podrían beneficiarse de datos más limpios y estandarizados para el análisis. La aplicación de estos modelos podría ser valiosa en el ámbito de la conservación y la investigación ecológica, donde la precisión y calidad de los datos son esenciales para el monitoreo efectivo de la biodiversidad.

Una señal libre de ruido permite además el desarrollo de un algoritmo de identificación de eventos que detecte y marque los puntos de corte en la señal. Esta identificación precisa de eventos clave en la señal limpia posibilitaría realizar el corte correspondiente en la señal original, maximizando la integridad de los datos y proporcionando segmentos específicos de alta relevancia para el análisis posterior.

Se propone aprovechar la última capa de los modelos de clasificación como un embedding para representar la señal en un espacio de características. Plantear el problema desde una perspectiva no supervisada, comparando los embeddings de audios procesados con reducción de ruido con los originales, podría revelar patrones latentes en las señales. Esta estrategia contribuiría a desarrollar un análisis robusto y eficiente en escenarios donde la supervisión directa de la clasificación no es posible o resulta limitada.

Finalmente, Un aspecto clave a explorar en futuras investigaciones es la evaluación del desempeño de BirdNET en la clasificación de especies de aves en Colombia, dado que su entrenamiento se ha basado principalmente en especies de América del Norte y Europa. Colombia, al ser uno de los países con mayor diversidad aviar del mundo y poseer numerosas especies endémicas, presenta un desafío significativo para la generalización del modelo. En este sentido, sería relevante comparar su rendimiento con bases de datos estandarizadas de

aves y anfibios de otras regiones, permitiendo una evaluación más robusta de su capacidad para clasificar especies fuera de su dominio de entrenamiento original.

Asimismo, se sugiere realizar experimentos complementarios utilizando bases de datos especializadas en aves y anfibios de América del Norte y Europa, con el objetivo de analizar el impacto de las técnicas de reducción de ruido en modelos bioacústicos en un contexto más controlado. Esta validación cruzada permitiría establecer con mayor claridad si las mejoras observadas en la clasificación de especies se deben a la efectividad del método de reducción de ruido o a la composición del conjunto de datos. Además, una discusión más detallada sobre la aplicabilidad de BirdNET en el contexto sudamericano.contribuiría a una mejor interpretación de los resultados y a futuras mejoras en los modelos bioacústicos para la biodiversidad.

XI. Anexos

XI-A. *Interfaz de Usuario y Funcionamiento de la Aplicación*

La aplicación desarrollada utiliza `Gradio` para proporcionar una interfaz gráfica de usuario intuitiva para el procesamiento de señales bioacústicas, específicamente en el ámbito del denoising y clasificación de audio en entornos ruidosos. La interfaz permite al usuario cargar un archivo de audio, seleccionar un modelo de clasificación (*BirdNet* o *AnuraSet*), y aplicar métodos de reducción de ruido basados en wavelets para mejorar la calidad de la señal.

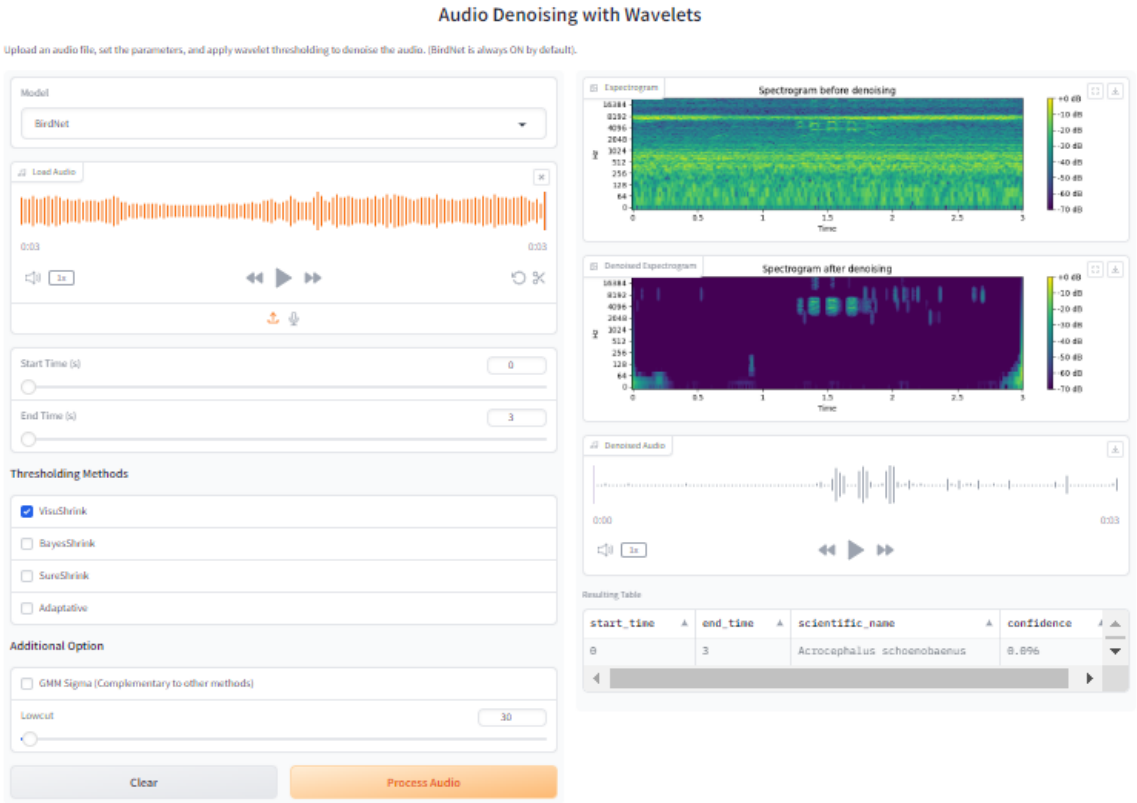


FIG. 32: Interfaz gráfica de la aplicación para denoising de audio con Wavelets en Gradio

La sección principal de la interfaz incluye los siguientes componentes:

- **Selección del Modelo:** En el menú desplegable, el usuario puede elegir entre dos modelos de clasificación de audio: *BirdNet*, diseñado para identificar vocalizaciones de aves, y *AnuraSet*, enfocado en la clasificación de sonidos de anuros. Cada modelo tiene su propio conjunto de wavelets optimizados para el procesamiento de señales específicas de cada grupo.

- **Carga de Audio:** La aplicación permite cargar archivos de audio en formato `.wav`. Una vez cargado, el usuario puede visualizar la forma de onda y seleccionar un intervalo temporal específico (inicio y fin) para analizar una sección concreta del audio.

- **Métodos de Umbralización:** Se proporcionan varios métodos de reducción de ruido basados en wavelets, como *VisuShrink*, *BayesShrink*, *SureShrink* y *Adaptative*. Estos métodos permiten ajustar la intensidad de reducción de ruido según la naturaleza del audio y el nivel de ruido presente.

- **Opciones Adicionales:** La opción de *GMM Sigma* permite complementar otros métodos de reducción de ruido mediante un ajuste basado en el modelo de mezcla gaussiana, lo cual mejora la precisión en ciertos casos de señales contaminadas. Adicionalmente, el usuario puede ajustar un filtro de paso bajo (parámetro `Lowcut`) para filtrar frecuencias no deseadas en el espectro.

- **Visualización de Espectrogramas:** Después de procesar el audio, se muestran dos espectrogramas: uno antes y otro después del denoising. Esto permite al usuario observar visualmente el efecto de la reducción de ruido en la señal bioacústica.

- **Audio Procesado y Tabla de Resultados:** La interfaz proporciona el audio denoised junto con una tabla que muestra los resultados de clasificación, que incluyen la especie identificada (`scientific_name`) y el nivel de confianza de la predicción (`confidence`). Este resultado es particularmente útil para la investigación en entornos de conservación y estudios de biodiversidad.

El flujo de trabajo se resume en la selección del modelo, ajuste de parámetros de denoising y ejecución del proceso de clasificación. El audio es preprocesado mediante la técnica de wavelets seleccionada y luego clasificado usando el modelo escogido. La interfaz permite descargar tanto el audio denoised como los resultados en formato de tabla para un análisis posterior.

XI-B. Procedimiento para Solicitud de Acceso bajo Confidencialidad

En caso de interés en acceder a la aplicación funcional o al repositorio asociado a este proyecto, se podrá solicitar el enlace correspondiente al autor mediante una comunicación formal. Toda solicitud deberá garantizar el cumplimiento de un acuerdo de confidencialidad, asegurando que el contenido no será divulgado, replicado ni distribuido sin autorización expresa. El acceso se concederá únicamente bajo el compromiso explícito de respetar los derechos de autor y de propiedad intelectual relacionados con este trabajo.

XII. Referencias

[1] S. Kahl, C. M. Wood, M. Eibl y H. Klinck, «BirdNET: A deep learning solution for avian diversity monitoring», *Ecological Informatics*, vol. 61, pág. 101 236, 2021, 1. Desarrollo de BirdNET, un modelo de aprendizaje profundo para identificar especies aviares a partir de datos acústicos. 2. Evaluación del rendimiento de BirdNET mediante conjuntos de datos independientes. DOI: 10.1016/j.ecoinf.2021.101236. dirección: https://github.com/kahst/BirdNET-Analyzer.

[2] M. A. Acevedo, C. J. Corrada-Bravo, H. Corrada-Bravo, L. J. Villanueva-Rivera y T. M. Aide, «Automated classification of bird and amphibian calls using machine learning: A comparison of methods», *Ecological Informatics*, vol. 4, n.º 4, págs. 206-214, 2009, 1. Comparación de algoritmos de aprendizaje automático para clasificación de llamadas de aves y ranas. 2. Evaluación del impacto de variables de llamadas estándar y extendidas en la precisión de clasificación., ISSN: 1574-9541. DOI: 10.1016/j.ecoinf.2009.06.005. dirección: https://doi.org/10.1016/j.ecoinf.2009.06.005.

[3] J. S. Cañas, M. P. Toro-Gómez, L. S. M. Sugai et al., «A dataset for benchmarking Neotropical anuran calls identification in passive acoustic monitoring», *Scientific Data*, vol. 10, págs. 1-12, 2023, 1. Desarrollo de un conjunto de datos anotados para llamadas de anuros en las regiones neotropicales. 2. Implementación de modelos de aprendizaje profundo para la identificación automatizada de llamadas de anuros. DOI: 10.1038/s41597-023-02666-2. dirección: https://github.com/soundclim/anuraset.

[4] e. a. Alonso J.B., «Automatic anuran identification using noise removal and audio activity detection», *Expert Systems with Applications*, vol. 72, págs. 83-92, 2017, Identificación automática de anuros con eliminación de ruido y detección de actividad acústica. DOI: 10.1016/j.eswa.2016.12.012.

[5] M. A. Acevedo y L. J. Villanueva-Rivera, «Using automated digital recording systems as effective tools for the monitoring of birds and amphibians», *Wildlife Society Bulletin*, vol. 34, n.º 1, págs. 211-214, 2006, 1. Evaluación de la eficacia de los sistemas de grabación digital automatizados (ADRS). 2. Comparación entre métodos tradicionales y ADRS para el monitoreo de aves y anfibios. DOI: 10.2193/0091-7648(2006)34[211:UADRSA]2.0.CO;2.

[6] A. Digby, M. Towsey, B. Bell y P. Teal, «A practical comparison of manual and autonomous methods for acoustic monitoring», *Methods in Ecology and Evolution*, vol. 4, n.º 7, págs. 675-683, 2013, 1. Comparación de métodos autónomos y manuales para monitoreo acústico. 2. Evaluación de la eficiencia y precisión de los métodos de detección de llamadas. DOI: 10.1111/2041-210X.12060.

[7] N. Priyadarshani, S. Marsland e I. Castro, «Automated birdsong recognition in complex acoustic environments: a review», *Journal of Avian Biology*, vol. 49, n.º 5,

págs. 1-27, 2018, 1. Reconocimiento automatizado de cantos de aves en entornos acústicos complejos. 2. Desafíos y mejoras en la eficacia y usabilidad de los sistemas de identificación automatizados de cantos de aves., ISSN: 0908-8857. DOI: 10.1111/jav.01447.

[8] A. Gómez, J. P. Ugarte y D. M. Murillo, «Bioacoustic Signals Denoising Using the Undecimated Discrete Wavelet Transform», *Communications in Computer and Information Science*, vol. 916, págs. 300-308, 2018, 1. Método de reconstrucción: umbral suave en dominio de wavelets. 2. Evaluación de eficacia y comparación con otros métodos estadísticos. DOI: 10.1007/978-3-030-00353-1_27.

[9] D. Martínez-Medina, O. Acevedo-Charry, S. Medellín-Becerra et al., «Estado, desarrollo y tendencias de los estudios en acústica de la fauna en Colombia», *Biota Colombiana*, vol. 22, n.º 1, págs. 7-20, 2021, 1. Análisis del estado de la investigación en bioacústica en Colombia. 2. Identificación de vacíos en el conocimiento y sugerencias para futuras investigaciones. DOI: 10.21068/c2021.v22n01a01.

[10] Z. Zhao, S. h. Zhang, Z. y. Xu et al., «Automated bird acoustic event detection and robust species classification», *Ecological Informatics*, vol. 79, NA, 2017, 1. Desarrollo de métodos automáticos para análisis de grabaciones de campo. 2. Mejora en la clasificación de especies de aves mediante modelos de reconocimiento de patrones. DOI: 10.1016/j.ecoinf.2017.04.003.

[11] L. F. Chaparro, *Signal and System Using MATLAB*. Department of Electrical y Computer Engineering, University of Pittsburgh, Year.

[12] L. Wasserman, «All of Nonparametric Statistics», en *Springer Texts in Statistics*, Springer, Year, cap. 9.

[13] H. Nyquist, «Certain Topics in Telegraph Transmission Theory», *Transactions of the A. I. E. E.*, vol. 248, NA, 1928, Teoría de sistemas de telegrafía; características de transmisión ideales. Se discuten criterios de transmisión perfecta, métodos para superar la ineficiencia de la telegrafía de portadora, análisis de formas de onda y efectos de interferencia y condiciones no ideales. DOI: 10.1109/5.989875.

[14] T. Napier, E. Ahn, S. Allen-Ankins, L. Schwarzkopf e I. Lee, «Advancements in preprocessing, detection and classification techniques for ecoacoustic data: A comprehensive review for large-scale Passive Acoustic Monitoring», *Expert Systems With Applications*, vol. 252, pág. 124 220, 2024, Revisión exhaustiva de las técnicas de preprocesamiento, detección y clasificación para el monitoreo acústico pasivo a gran escala. DOI: 10.1016/j.eswa.2024.124220.

[15] G. Huang, D. Ferreira, P. Gomes et al., «Intelligent feature extraction and classification of anuran vocalizations», *Applied Soft Computing*, vol. 19, págs. 165-179, 2014, 1. Uso de wavelets para la extracción de características de vocalizaciones de anuros. 2. Mejora en la clasificación de señales bioacústicas ruidosas. DOI: 10.1016/j.asoc.2014.01.037.

[16] Y. Ren, M. T. Johnson y J. Tao, «Perceptually motivated wavelet packet transform for bioacoustic signal enhancement», *The Journal of the Acoustical Society of America*, vol. 124, n.º 1, págs. 316-327, 2008, 1. Uso de wavelets para mejorar señales bioacústicas en ambientes ruidosos. 2. Comparación con métodos tradicionales como la filtración de Wiener. DOI: 10.1121/1.2936465.

[17] Y. Ren, M. Johnson y J. Tao, «Perceptually motivated wavelet packet transform for bioacoustic signal enhancement», *The Journal of the Acoustical Society of America*, vol. 124, n.º 1, págs. 316-327, 2008, Uso de wavelets para mejorar señales bioacústicas en ambientes ruidosos. DOI: 10.1121/1.2936465.

[18] D. L. Donoho, «De-Noising by Soft-Thresholding», *IEEE Transactions on Information Theory*, vol. 41, n.º 3, 1995, 1. Modelo de reconstrucción de Donoho y Johnstone (1994), basado en umbral suave en el dominio de wavelets. DOI: 10.1109/18.382009.

[19] H. A. Chipman, E. D. Kolaczyk y R. E. McCulloch, «Adaptive Bayesian Wavelet Shrinkage», *Journal of the American Statistical Association*, vol. 92, n.º 440, págs. 1413-1421, 1997, 1. Propuesta de un método bayesiano adaptativo para la contracción de coeficientes wavelet. 2. Comparación con otros métodos de contracción y evaluación de bandas de incertidumbre. DOI: 10.1080/01621459.1997.10473662.

[20] G. Chen y T. Bui, «Multiwavelets denoising using neighboring coefficients», *IEEE Signal Processing Letters*, vol. 10, n.º 7, págs. 211-214, 2003, Propuesta de denoising mediante multiwavelets usando coeficientes vecinos. DOI: 10.1109/LSP.2003.813211.

[21] Z. Varady, «Adaptive wavelet denoising of biosignals using statistical models», *IEEE Transactions on Biomedical Engineering*, vol. 48, n.º 5, págs. 592-596, 2001, Métodos adaptativos de denoising mediante wavelets para señales biomédicas utilizando modelos estadísticos. DOI: 10.1109/10.929628.

[22] V. Srivastava, R. Srivastava y V. Srivastava, «Optimal level selection criterion for wavelet denoising of audio signals», *Applied Acoustics*, vol. 113, págs. 7-15, 2016, Discusión sobre criterios de selección de niveles de descomposición óptimos para denoising en señales de audio. DOI: 10.1016/j.apacoust.2016.05.006.

[23] C. E. Shannon, «A Mathematical Theory of Communication», *Bell System Technical Journal*, vol. 27, págs. 379-423, 1948, Fundamento de la teoría de la información y la entropía como medida de incertidumbre. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[24] W. Ngui, M. S. H. Leong, M. Lim y A. M. Abdelrhman, «Wavelet analysis: Mother wavelet selection methods», *Applied Mechanics and Materials*, vol. 393, págs. 953-958, 2013, 1. Métodos de selección de madre wavelet. 2. Aplicaciones wavelet en análisis de señales. DOI: 10.4028/www.scientific.net/AMM.393.953.

[25] H.-f. Ma, J.-w. Dang y X. Liu, «Research of the Optimal Wavelet Selection on Entropy Function», *Lecture Notes in Electrical Engineering*, vol. 2, págs. 35-42, 2012, 1. Selección óptima de wavelets. 2. Mejora de la denoización de señales de voz. DOI: 10.1007/978-3-642-31003-4_5.

[26] F. Chollet, *Deep Learning with Python*. Manning Publications Company, 2017.

[27] G. James, D. Witten, T. Hastie y R. Tibshirani, *An Introduction to Statistical Learning with Applications in Python*. Springer, 2013.

[28] S. C. Salazar, *Bioacoustic Signal Refine Classifier*, GitHub repository, 2024. dirección: https://github.com/SebastianCarvalhoSalazar/BioacousticSignalRefineClassifier.