

A Corpus of Bird Sounds from Quindío and its Application for Passive Acoustic Monitoring through Neural Networks

Helver Novoa Mendoza , Alexandra Rengifo , Emilio Granell , and Fáber Danilo Giraldo 

Abstract—The biodiversity of a region is an invaluable asset that requires continuous efforts for its preservation and study. The Quindío region in Colombia, in particular, is renowned for its richness in native birdlife, which enhances its natural landscape and significantly contributes to the country's biological diversity. Observing and studying these species is essential for the conservation and understanding of the region's ecology. In this context, this paper proposes a sound corpus supported by machine learning (ML) processing. The dataset includes audio recordings of 170 bird species, representing approximately 30% of the bird species identified in Quindío. Additionally, the corpus contains recordings of human voices, ambient silences, and background noises. For signal processing, we employed the sliding window feature extraction technique to analyze and classify bird sounds. Three neural networks were trained to evaluate the corpus. The first was a convolutional neural network (CNN). Based on the results of this initial network, two additional networks were trained: another CNN and a network based on the transformer architecture. These networks were trained with categories that achieved an F1-score metric of 0.30 or higher in the first CNN. The results indicated precision levels of 0.55, 0.53, and 0.65, respectively. The transformer-based network demonstrated superior performance in classifying the native bird sounds of Quindío. A proof of concept was conducted using audio recordings of the Saffron Finch (*Sicalis flaveola*), achieving an accuracy of 65.74%. These findings establish a baseline for future research in the field of bird sound classification, thereby promoting the conservation of the region's avifauna.

Link to graphical and video abstracts, and to code:
<https://latamt.ieeer9.org/index.php/transactions/article/view/9191>

Index Terms—Passive acoustic monitoring, bird sounds, sound corpus, avifauna, feature extraction.

I. INTRODUCCIÓN

EL monitoreo de paisajes acústicos se ha consolidado como una herramienta no invasiva clave para obtener información biológica, sin la necesidad de capturar o sacrificar individuos. A través del uso de software especializado como

The associate editor coordinating the review of this manuscript and approving it for publication was Ruth Aguilar (*Corresponding author: Fáber D. Giraldo*).

This work was supported by Universidad del Quindío (100016837) under Grant (1194) and Sistema General de Regalías under Grant (2018000040015).

Fáber D. Giraldo-Velásquez, A. Rengifo, and H. N. Mendoza are with the Universidad del Quindío, Armenia, Colombia (e-mails: fdgiraldo@uniquindio.edu.co, alexandra.rengifor@uqvirtual.edu.co, and hanovoam@uqvirtual.edu.co).

E. Granell is with the Universitat Politècnica de Valencia, Valencia, Spain (e-mail: egranell@upv.es).

Kaleidoscope¹, se pueden identificar especies de animales, incluidas aves y murciélagos, a partir de grabaciones de sonido. Este enfoque ha ganado relevancia por su bajo costo y capacidad para evaluar la riqueza y complejidad de los ecosistemas, ofreciendo una amplia cobertura espacial y temporal sin perturbar el entorno natural.

El procesamiento eficiente de grandes conjuntos de datos de audio para extraer información precisa sobre la diversidad de especies ha sido un desafío. Sin embargo, proyectos como BirdNET [1], desarrollado por el laboratorio de ornitología de la Universidad de Cornell, han sido pioneros en el uso de técnicas de aprendizaje automático para la clasificación de aves. Su aplicación basada en redes neuronales convolucionales permite identificar más de 900 especies de aves en América del Norte y Europa. De forma similar, la aplicación Merlin Bird ID² actúa como una guía de campo personalizada para la identificación de aves en diferentes regiones del mundo. Estas herramientas demuestran el potencial de la tecnología en la identificación de la avifauna y subrayan la importancia de métodos automáticos para su monitoreo.

La bioacústica, disciplina que combina la biología y la acústica, ha tenido un gran desarrollo gracias a la incorporación de modelos de aprendizaje automático y redes neuronales. Los estudios bioacústicos permiten investigar la comunicación animal a través de señales acústicas, con aplicaciones prácticas en la conservación y monitoreo de la biodiversidad. Además, esta disciplina ha demostrado su utilidad en la valoración de la biodiversidad mediante la interpretación del "paisaje sonoro", compuesto por biofonías, geofonías y antropfonías [2]. La creciente disponibilidad de dispositivos de grabación autónoma y métodos de análisis avanzados está permitiendo un análisis más detallado y continuo de estos paisajes sonoros, lo que facilita la comprensión de la dinámica de los ecosistemas y la interacción de las especies con su entorno acústico [2].

En este sentido, el estudio realizado en la Reserva Natural Los Yátaros, en Boyacá, Colombia, durante el primer semestre de 2020 [3], muestra un ejemplo práctico de la bioacústica a través del muestreo acústico pasivo. Se utilizaron seis grabadoras automáticas AudioMoth, dos en cada sitio de muestreo, programadas para capturar un minuto de audio cada 30 minutos a lo largo de todo el día en el espectro audible (20 Hz - 20 kHz) y durante las noches en ultrasonido (hasta 192 kHz). Este enfoque de muestreo permite recopilar

¹<https://www.wildlifeacoustics.com/>

²Disponible en <https://merlin.allaboutbirds.org>

datos continuos y detallados del paisaje sonoro, brindando información valiosa sobre la diversidad acústica y los patrones temporales de la vida silvestre en el área estudiada.

En este trabajo, nos centramos en la avifauna del departamento del Quindío, Colombia, una región de alta biodiversidad. A partir de los sonidos recopilados en esta región, construimos un corpus acústico y entrenamos modelos de redes neuronales, incluyendo redes convolucionales y una arquitectura basada en Transformers, para evaluar la diversidad acústica. A través de este enfoque, buscamos contribuir al monitoreo y conservación de la biodiversidad avifaunística, proporcionando una herramienta precisa y eficaz para la clasificación de especies a partir de sus sonidos. Además, se exploran métricas como el Rango Recíproco Medio (MRR) para validar el desempeño del modelo propuesto.

El resto del artículo se estructura de la siguiente forma: la sección (Sección II) presenta una revisión de trabajos previos; nuestra propuesta se detalla en la Sección III; el entorno de experimentación se describe en la Sección IV; los experimentos y los resultados obtenidos se reportan en la Sección V; una prueba de concepto con audios de la especie Saffron Finch (*Sicalis flaveola*) en la sección VI; las conclusiones y las líneas de trabajo futuro se trazan en la Sección VII. La información necesaria para reproducir este trabajo se detalla en la Sección VIII.

II. TRABAJOS PREVIOS

Los primeros trabajos de monitoreo acústico pasivo, algunos ya descritos, representaron un hito en la investigación científica. Estos estudios pioneros se centraron en la utilización de micrófonos estratégicamente ubicados para capturar y analizar los sonidos presentes en un entorno sin la intervención activa del observador. Mediante este enfoque novedoso, se logró obtener valiosa información sobre la diversidad y distribución de especies animales, así como la identificación de patrones acústicos y comportamientos específicos. Estos primeros avances en el monitoreo acústico pasivo allanaron el camino para el desarrollo de técnicas más sofisticadas, contribuyendo así a la comprensión y conservación de los ecosistemas.

A. Avances en Clasificación Acústica Automática

En el trabajo de J. Salamon *et al.* [4] se emplearon técnicas de clasificación de vanguardia para identificar especies de aves a partir de sus llamadas de vuelo, para esto emplearon las características de los Coeficientes Cepstrales en las Frecuencias de Mel (MFCC por sus siglas en inglés de *Mel Frequency Cepstral Coefficients*). En el estudio, se compara un enfoque de “aprendizaje superficial” basado en el aprendizaje no supervisado a partir de los MFCC, con una red neuronal convolucional profunda junto con técnicas de aumento de datos. Los resultados mostraron que ambos enfoques obtienen resultados comparables en un conjunto de datos considerable que incluye 5428 llamadas de vuelo de 43 especies diferentes. Al fusionar los modelos de aprendizaje no supervisado y supervisado, lograron una precisión del 96% en la clasificación de aves. Estos hallazgos resaltan el potencial de las técnicas

de aprendizaje automático para la identificación precisa de especies de aves mediante sus llamadas de vuelo, lo que podría beneficiar significativamente la investigación y conservación de las aves.

Por su parte, la investigación de F.J. Bravo *et al.* [5] aborda una problemática común en el desarrollo de modelos de aprendizaje supervisado: la escasez de datos etiquetados. En este contexto, los investigadores exploraron el uso de *SincNet*, una arquitectura eficiente de red neuronal que aprende directamente de las formas de onda sin procesar mediante filtros sincronizados. Los resultados obtenidos con SincNet revelaron que la red neuronal converge rápidamente y logra precisiones superiores al 65% incluso con conjuntos de datos limitados. Los autores concluyen que el enfoque de aprendizaje directo de las formas de onda sin procesar permite que el algoritmo seleccione automáticamente los elementos sonoros más relevantes para la tarea, evitando así la necesidad de utilizar técnicas de extracción de características y reduciendo posibles sesgos. Este enfoque prometedor muestra cómo las técnicas de procesamiento de señales pueden mejorar la eficiencia y precisión de los modelos de aprendizaje automático en condiciones de datos limitados.

En el mismo sentido, en el trabajo de I. Kiskin *et al.* [6] se aborda el desafío de la detección de eventos acústicos en un contexto de escasez de datos utilizando técnicas de aprendizaje profundo. Los investigadores aplican transformaciones wavelet, que ofrecen ventajas sobre la transformada de Fourier, para entrenar redes neuronales convolucionales. Los resultados muestran un rendimiento superior en comparación con los clasificadores convencionales. El modelo desarrollado tiene como objetivo detectar la presencia de mosquitos y clasificar especies de aves. Este enfoque resalta la capacidad de las técnicas de aprendizaje profundo para abordar problemas de detección en condiciones de datos limitados y la importancia de las transformaciones wavelet en el análisis acústico.

La propuesta presentada en [7], caracterizada por ser una de las primeras en utilizar técnicas automatizadas para el reconocimiento bioacústico de plagas de insectos específicas, se enfocó en capturar y reconocer automáticamente las emisiones acústicas resultantes de comportamientos típicos de las plagas objetivo, como la locomoción y la alimentación. Después de adquirir los sonidos, los amplificaron, los filtraron, los parametrizaron y los clasificaron mediante métodos avanzados de aprendizaje automático en una computadora portátil. Al igual que en los trabajos previos, se emplearon los MFCC como el tipo de información utilizado. Este enfoque demostró la viabilidad de utilizar el análisis acústico automatizado para la detección y monitoreo de plagas de insectos.

La aplicación de redes neuronales convolucionales (CNN, por sus siglas en inglés de *Convolutional Neural Networks*) y de modelos con Transformadores (*transformers*) para el reconocimiento de sonidos de aves, fundamentan un trabajo [8] cuyo objetivo principal es estudiar las poblaciones de aves en peligro de extinción a través del monitoreo bioacústico. En su trabajo se describen los desafíos que enfrentan los investigadores al monitorear las poblaciones de aves y cómo el monitoreo bioacústico ayuda a superarlos. También discuten las técnicas utilizadas para procesar grandes conjuntos de datos

bioacústicos utilizando aprendizaje automático. Concluyen con una discusión sobre los beneficios potenciales del uso del aprendizaje automático para el reconocimiento de sonidos de aves y cómo esto podría impactar los esfuerzos de conservación.

Por su parte, en [9] se explora el uso de la auto-atención en la identificación de llamadas de aves en paisajes sonoros, utilizando los avances recientes en Aprendizaje Automático y Aprendizaje Profundo. Los resultados muestran que los modelos basados en Transformadores superan a los modelos basados en CNN, lo que proporcionaría una vía prometedora para el reconocimiento automático de aves en entornos diversos. Por último, en [10] se aborda el uso de los Transformadores STFT (siglas de *Short-time Fourier transform*, Transformada de Fourier de Tiempo Corto) para el reconocimiento de cantos de aves. En particular, describen un nuevo modelo basado en Transformadores que trabaja directamente con secuencias de STFT. El modelo se basa en los transformadores de visión (ViT) y se adapta a la clasificación de espectrogramas.

B. Análisis Acústico en Ecosistemas

Con respecto al estudio de los ecosistemas, en el trabajo presentado en [11] se propone una metodología para analizar la dinámica de éstos mediante el monitoreo acústico. Para ello, se integran varios índices acústicos que buscan medir la calidad de los hábitats a partir de grabaciones. Los índices acústicos, denominados alfa, representan diferentes atributos de un hábitat, como la uniformidad, la riqueza y la heterogeneidad. Los investigadores utilizaron métodos de clasificación como redes multicapa y máquinas de soporte vectorial de clase única para caracterizar e identificar prototipos de hábitat predefinidos. Los resultados mostraron una precisión de 0.89 ± 0.01 en la clasificación de tres hábitats con diferentes grados de perturbación. Además, se realizó un experimento adicional para validar la metodología a una resolución más fina, logrando identificar tres tipos de bosques. La metodología propuesta demostró su eficacia para caracterizar y clasificar hábitats utilizando el monitoreo acústico.

El monitoreo acústico pasivo (PAM, por las siglas en inglés de *Passive Acoustic Monitoring*) ha experimentado un rápido crecimiento en la investigación ecológica, siguiendo la tendencia global hacia la recopilación automatizada de datos y el análisis masivo de ellos (*big data*). En el estudio de [12] se llevó a cabo una revisión sistemática del PAM terrestre, examinando 460 artículos publicados en 122 revistas entre 1992 y 2018. Durante este período, se observó un aumento significativo en las publicaciones relacionadas con el PAM, que abarcaron tres fases de desarrollo: establecimiento, expansión y consolidación. En general, la investigación se ha centrado principalmente en murciélagos (50%) y en regiones templadas del norte (65%), con un enfoque en los patrones de actividad nocturna (25%). Aunque el uso de registradores no programables (61%) y el análisis acústico manual (58%) ha sido común, se espera una mayor diversificación de aplicaciones en el futuro.

El estudio de [13] se enfoca en el uso del monitoreo acústico pasivo para evaluar poblaciones de primates salvajes

en el bosque del Parque Nacional Tai en Costa de Marfil. Los investigadores exploraron el uso de un algoritmo personalizado para la detección automática de múltiples especies de primates, con el objetivo de obtener estimaciones precisas de su presencia a partir de datos acústicos. Utilizando grabaciones continuas de la selva tropical y unidades de grabación autónomas (ARU, por sus siglas en inglés de *Autonomous Recording Unit*), se detectaron y clasificaron señales de sonido específicas de diferentes especies. A través de un enfoque de modelado, se investigó la necesidad de intervención manual por parte de expertos humanos para mejorar las probabilidades derivadas de los datos de ARU. Este estudio se destaca por la combinación de métodos automatizados y la experiencia humana en la identificación precisa de especies en entornos naturales.

Enfocado en los efectos relacionados con el clima, la propuesta de [14] resalta la necesidad de comprender cómo el cambio global afecta la vida silvestre y los ecosistemas. Identifica en el sonido un componente fundamental de cualquier hábitat y destaca el valor de las grabaciones de sonido realizadas en el campo como una fuente rica de información ecológica. Estas grabaciones pueden proporcionar datos importantes sobre la abundancia, la distribución y el comportamiento de los animales que emiten vocalizaciones en un área determinada. En este contexto, los sensores acústicos están siendo ampliamente utilizados en entornos ecológicos y de conservación para monitorear las poblaciones de animales, su comportamiento y sus respuestas ante los cambios ambientales. Este enfoque innovador brinda una herramienta invaluable para comprender y proteger mejor la biodiversidad y los ecosistemas en un mundo en constante transformación.

C. Trabajos de Bioacústica en Colombia

En el contexto nacional, el estudio presentado en [15] describe una metodología no supervisada para la identificación automática de anuros en Colombia. Este enfoque se basa en el uso de un clasificador difuso de los MFCC. Lo interesante de esta metodología es que puede detectar especies que no fueron presentadas durante la etapa de entrenamiento, incluso si pertenecen a poblaciones diferentes. Además, las similitudes en las llamadas permiten determinar correlaciones entre especies del mismo género. Los autores realizaron pruebas utilizando dos conjuntos de datos diferentes con especies del noreste de Colombia, obteniendo precisiones de entre el 99.38% y el 100% en todas las especies. Los autores destacan la efectividad de la metodología propuesta para la identificación precisa de anuros en diferentes conjuntos de datos.

En [16] se destaca la labor de un grupo de científicas que investigan los sonidos en el Golfo de Tribugá, en el Pacífico sur de Colombia, con el objetivo de conservar este ecosistema. Su trabajo se centra en comprender cómo se comunican los animales bajo el agua y en estudiar las dinámicas temporales y espaciales de sus interacciones. Además, buscan evaluar el impacto humano en el paisaje acústico y proponer medidas de gestión del ruido para evitar perturbaciones en los ciclos biológicos naturales de los animales. Con esta investigación se

busca promover la conservación de la vida marina y garantizar la sostenibilidad de este importante hábitat acuático.

Por último, en [17] se resaltan los beneficios de la bioacústica digital en el descubrimiento de sonidos naturales ocultos, especialmente aquellos que se encuentran en frecuencias ultrasónicas o infrasónicas, más allá del rango auditivo humano. Se mencionan estudios sobre las crías de tortugas, que anteriormente se creía que eran mudas. Gracias a la bioacústica, se ha descubierto que estas crías emiten sonidos mientras aún están dentro de sus huevos, lo cual sugiere que estos sonidos tienen la función de coordinar el momento de su nacimiento. Estos hallazgos demuestran cómo la bioacústica permite adentrarse en el mundo sonoro de la naturaleza y revelar aspectos desconocidos de la comunicación animal.

III. PROPUESTA

El aprendizaje profundo ha revolucionado el campo de la inteligencia artificial, permitiendo avances significativos en el procesamiento y análisis de datos complejos. Los modelos de aprendizaje profundo, en particular las redes neuronales, son capaces de aprender representaciones jerárquicas de los datos, lo que los hace adecuados para tareas de clasificación y reconocimiento de patrones.

En este sentido, la consolidación de un corpus de sonidos de aves es un componente esencial para la implementación del monitoreo acústico pasivo. Para lograr esto, la técnica de ventanas deslizantes permite la extracción de características relevantes de los segmentos de audio recopilados. El entrenamiento de una red convolucional que evalúa la calidad del corpus en función de parámetros específicos. A partir de esta red se seleccionan las categorías con un rendimiento en la métrica F1-score igual o superior a 0.30 para varias.

Para continuar refinando el proceso de clasificación, se entrenan dos redes adicionales: una red convolucional y una Transformer. Estas redes se centran en las categorías que muestran un rendimiento adecuado en la primera red convolucional. La comparación de los resultados obtenidos con las tres redes arroja una visión de su capacidad para identificar y clasificar las especies de aves en el corpus de sonidos. Este enfoque integral permite una evaluación rigurosa del corpus y destaca la importancia de una selección adecuada de las técnicas de procesamiento de señales y aprendizaje profundo en el monitoreo acústico de la biodiversidad aviar.

A. Corpus de Sonidos de Aves del Quindío

El corpus de sonidos de aves utilizado fue construido a partir de una exhaustiva recopilación proporcionada por Aicrowd. Este corpus incluye un total de 170 especies de aves identificadas en el departamento del Quindío. Cada especie cuenta con múltiples grabaciones, capturando una amplia variedad de vocalizaciones y comportamientos acústicos. Para garantizar la representatividad y diversidad del corpus, también se han incluido audios de voces humanas, silencios y ruidos ambientales. <https://www.aicrowd.com>.

Con un total de 56725 audios y una media de 327.89 grabaciones por cada una de las 170 especies identificadas en

el departamento del Quindío, este corpus ofrece una importante fuente de información para el desarrollo de sistemas de reconocimiento de vocalizaciones aviares. Estas grabaciones, que suman un total de 38.5 horas de sonido, han sido recopiladas de manera exhaustiva y han sido cuidadosamente seleccionadas para garantizar la calidad del conjunto de datos³. Se espera que este recurso sea un valioso punto de referencia para investigadores y entusiastas de las aves que buscan explorar y comprender la riqueza acústica de la fauna aviar del Quindío y contribuir al desarrollo de nuevas herramientas para su conservación y estudio.

Con el fin de facilitar el procesamiento y análisis de los audios, se realizó una segmentación en el corpus original. Se identificaron los periodos de tiempo en los que la frecuencia era inferior a los 12 KHz durante más de un segundo, y se utilizaron como puntos de corte para dividir los audios en segmentos más cortos. Esta segmentación permitió una mayor granularidad en el análisis y extracción de características, al tiempo que se conservó la información relevante de cada vocalización. Para la evaluación del corpus de sonidos de aves, se implementan modelos de redes neuronales convolucionales y Transformers. Estos modelos se constituyen en un indicador sobre la calidad de las grabaciones de aves, además de determinar sus capacidades para capturar patrones y características sutiles en los datos acústicos. La combinación de estas dos arquitecturas de aprendizaje profundo ofrece la posibilidad de identificar alternativas conducentes a mejorar la calidad de la clasificación y la eficiencia en el monitoreo acústico de la biodiversidad aviar.

B. Extracción de Características: Ventana Deslizante

La técnica de ventana deslizante es una estrategia utilizada para dividir una señal de sonido o una imagen en segmentos más pequeños y superpuestos. En el contexto se implementó la técnica de ventana deslizante utilizando Python, lo que permitió generar un total de 102 092 espectrogramas en escala de grises. El tamaño de cada espectrograma para las dos técnicas es el mismo: 224×224 píxeles.

La ventana deslizante implica el desplazamiento de una ventana rectangular a lo largo de la señal o la imagen original, capturando así diferentes segmentos en cada posición. Esta técnica es especialmente útil cuando se desea aumentar y a la vez conservar las características detalladas de una señal, o como en este caso, de los espectrogramas de aves. Al aplicar la ventana deslizante, se obtiene información más granular sobre las características temporales y frecuenciales presentes en los cantos de las aves. De esta manera, se facilita la identificación y clasificación precisa de las especies aviares mediante el análisis de segmentos más pequeños y específicos de sus vocalizaciones. La Fig. 1 muestra el número de ventanas resultantes sobre un audio de la especie *Accipiter bicolor* Bicolored Hawk.

³Para facilitar su acceso y uso, todos los audios se encuentran disponibles en formato .flac y pueden ser descargados a través del siguiente enlace: <https://zenodo.org/record/8322445>.

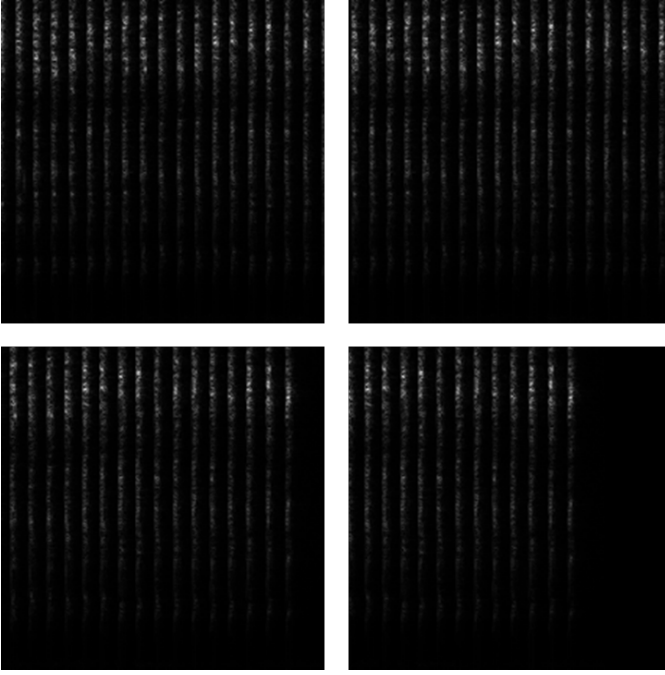


Fig. 1. Ventana deslizante: Espectrogramas de la especie *Accipiter bicolor* Bicolored Hawk.

IV. EXPERIMENTOS

En esta sección se describe el conjunto de datos para la técnica de extracción de características de ventana deslizante y las arquitecturas de las redes neuronales empleadas para su evaluación.

A. El Conjunto de Datos

Para la técnica de ventana deslizante, el conjunto de datos incluye espectrogramas de 170 especies de aves inventareadas. En este caso, la especie “*Panyptila cayennensis*” (Vencejo Rabo de Tijera Menor) presenta el menor número de espectrogramas generados, con un total de 38 muestras. Por otro lado, la especie con el mayor número de espectrogramas es “*Herpetotheres cachinnans*” (Halconcito Risueño), contando con 3 684 espectrogramas en el conjunto de datos.

B. Arquitectura de la CNN

La arquitectura de esta red convolucional se compone de varias capas que procesan imágenes de entrada de 224×224 píxeles en escala de grises. Comienza con una capa lambda que normaliza las imágenes de entrada. Luego, sigue con capas convolucionales y capas de max-pooling que extraen características de las imágenes. La primera capa convolucional tiene 30 filtros de 3×3 píxeles, y la segunda capa convolucional tiene 50 filtros de 5×5 píxeles, seguidos por capas de max-pooling que reducen la resolución de la imagen.

Después de estas capas de convolución y max-pooling, se aplica una capa de aplanamiento (flatten) que convierte los mapas de características 2D en un vector unidimensional. Luego, se agregan capas densas (fully connected) para realizar la clasificación. La primera capa densa tiene 500 neuronas,

seguida por una capa de dropout para regularizar el modelo. Luego, hay otra capa densa de 500 neuronas, nuevamente seguida por una capa de dropout. Finalmente, la última capa densa produce las predicciones de salida con 173 unidades, lo que sugiere que esta red está diseñada para la clasificación en 173 clases distintas.

TABLA I
ARQUITECTURA DE LA CNN

Capa	Tamaño de salida	Parámetros entrenables
Lambda	(None, 224, 224, 1)	0
Conv2D	(None, 222, 222, 30)	300
MaxPooling2D	(None, 111, 111, 30)	0
Conv2D	(None, 107, 107, 50)	37,550
MaxPooling2D	(None, 53, 53, 50)	0
Conv2D	(None, 49, 49, 50)	62,550
MaxPooling2D	(None, 24, 24, 50)	0
Flatten	(None, 28,800)	0
Dense	(None, 500)	14,400
Dropout	(None, 500)	0
Dense	(None, 500)	250,500
Dropout	(None, 500)	0
Dense	(None, 173)	86,673

En la Tabla I se resume cada capa de la red convolucional. En total, esta arquitectura tiene 14.838.073 parámetros entrenables, lo que implica un modelo con una capacidad significativa para la clasificación de imágenes. El uso de capas de dropout ayuda a prevenir el sobreajuste al regularizar el modelo durante el entrenamiento. Esta arquitectura puede ser adecuada para tareas de clasificación de imágenes en las que se requiere una extracción de características y clasificación avanzada.

C. Arquitectura de la Red Transformer

La arquitectura de la red presentada es un modelo Transformer, que se ha convertido en un estándar en tareas de procesamiento de lenguaje natural y visión por computadora. El modelo comienza con una capa de entrada que toma datos en formato de imágenes de 224×224 píxeles. Luego, se extraen parches de las imágenes y se codifican a través de la capa de PatchEncoder, que transforma estos parches en representaciones de baja dimensión. A continuación, se aplican múltiples capas de atención de múltiples cabezas (multi-head attention) y capas de normalización para capturar las relaciones a larga distancia entre los parches. Estas capas se conectan en serie mediante capas de atención y de tipo feedforward, y se emplea la técnica de dropout para regularizar el modelo y evitar el sobreajuste.

El modelo se repite en varias etapas, cada una compuesta por una secuencia de capas de atención y feedforward, lo que permite capturar información contextual a diferentes niveles de abstracción. Las capas de atención múltiple permiten que el modelo comprenda las relaciones espaciales y semánticas entre los parches de la imagen. Finalmente, después de múltiples repeticiones, se realiza una operación de aplanamiento (flatten) y se conecta a capas densas que culminan con una capa de salida que produce predicciones. En total, esta arquitectura contiene 36 580 517 de parámetros entrenables y en otros

contextos ha demostrado ser eficaz en tareas de visión por computadora y procesamiento de imágenes.

El componente innovador de esta propuesta radica en la aplicación de diferentes arquitecturas (CNN y Vision Transformer) para abordar el mismo problema. Esta estrategia permite aprovechar las fortalezas de cada técnica por separado, permitiendo un análisis más detallado y preciso de las vocalizaciones de aves. En lugar de fusionar las redes, se implementarán de manera paralela y se compararán los resultados obtenidos, lo que proporcionará una perspectiva más amplia sobre cuál de los modelos proporciona un rendimiento superior en la clasificación de vocalizaciones.

V. RESULTADOS

En esta sección se presentan los resultados de la evaluación del corpus de sonidos de aves, mediante las arquitecturas de redes convolucional y de Transformer propuestas.

A. Resultados de la CNN

A partir de la evaluación del modelo de red convolucional, se pueden extraer varias conclusiones significativas. En primer lugar, la función de pérdida muestra una tendencia a estabilizarse después de 40 épocas, lo que sugiere que el modelo alcanzó su capacidad máxima de aprendizaje en esta tarea particular. Además, la precisión del conjunto de validación se mantuvo en alrededor del 41.30%, lo que indica que el modelo es capaz de reconocer especies de aves con un nivel razonable de exactitud. La Fig. 2 y la Fig. 3 ilustran el progreso del entrenamiento de la red.

Sin embargo, es importante destacar que la precisión es del 55.44% en las muestras de entrenamiento, lo que puede indicar cierto grado de sobreajuste, ya que es significativamente mayor que la precisión en las muestras de validación. Esto sugiere que el modelo podría beneficiarse de una mayor regularización o un conjunto de datos de entrenamiento más diverso y amplio. Aunque esta red convolucional muestra una capacidad prometedora para el reconocimiento de sonidos de aves, es necesario aumentar el conjunto de datos y evaluar los criterios de selección del corpus de sonidos para mejorar su rendimiento en condiciones del mundo real.

La Tabla II y la Tabla III muestran el top 5 y el 5 más bajo para cada una de las categorías evaluadas. Estos resultados de la evaluación de la red convolucional entrenada muestran una variabilidad significativa en el rendimiento de la detección de especies de aves. Se observa que algunas especies, como “Herpetotheres cachinnans Laughing Falcon”, “Tapera naevia Striped Cuckoo”, y “Coccyzus erythrophthalmus Black-billed Cuckoo”, obtienen una alta precisión, superando el 80%, lo que indica una capacidad destacada para identificar estas especies en particular. Sin embargo, en el otro extremo del espectro, se encuentran especies como “Anhinga anhinga Anhinga”, “Anthracothorax nigricollis Black-throated Mango”, y “Nomonyx dominicus Masked Duck”, con puntajes de precisión muy bajos, lo que sugiere dificultades en su reconocimiento.

Lo anterior resalta los desafíos de la clasificación de especies de aves a través de la detección de sus sonidos.

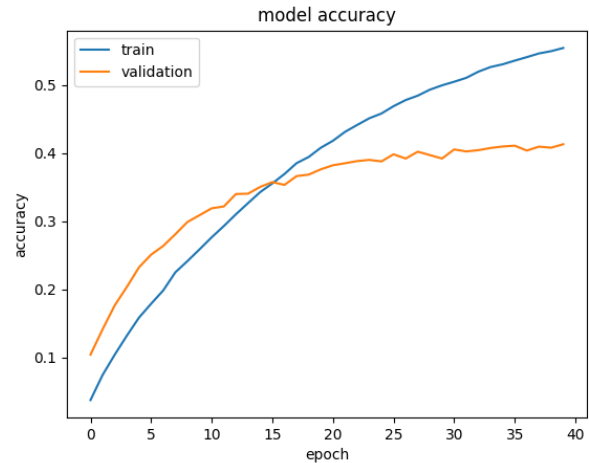


Fig. 2. CNN: Precisión para los conjuntos de entrenamiento y validación.

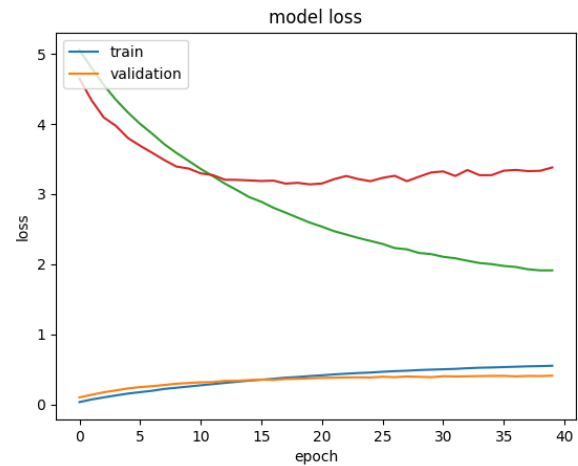


Fig. 3. CNN: Función de pérdida para los conjuntos de entrenamiento y validación.

Además, muestran la importancia de una mayor optimización del modelo y una mayor diversidad en el conjunto de datos para mejorar la capacidad de reconocimiento de las especies menos comunes. La aplicación de técnicas avanzadas de procesamiento de señales y el aumento de la cantidad de datos de entrenamiento pueden ser pasos futuros para mejorar la precisión en la identificación de especies de aves.

Basados en los resultados anteriores de la evaluación de la red convolucional, se seleccionaron las especies de aves cuyas métricas F1-score sean iguales o superiores a 0.30. El F1-score es una métrica que combina la precisión y el recall en una sola puntuación, lo que la hace especialmente útil cuando se tiene un desequilibrio entre las clases, como en este caso, donde algunas especies son menos comunes. Un F1-score de 0.30 o superior indica un equilibrio razonable entre la capacidad del modelo para identificar correctamente las especies y evitar falsos positivos. Esta elección permite centrarse en aquellas especies que el modelo puede identificar con un nivel aceptable de precisión y recall.

TABLA II
CNN. MÉTRICAS DE RENDIMIENTO POR ESPECIE. TOP 5

Especie	Precisión	Recall	F1-score	Support
Herpetoheres cacinans_Laughing Falcon	0.98	0.74	0.84	581
Tapera naevia_Striped Cuckoo	0.92	0.72	0.81	325
Coccyzus americanus_Yellow-billed Cuckoo	0.88	0.32	0.47	547
Columbina talpacoti_Ruddy Ground Dove	0.87	0.40	0.55	357
Vanellus chilensis_Southern Lapwing	0.85	0.66	0.75	338

TABLA III
CNN. MÉTRICAS DE RENDIMIENTO POR ESPECIE. 5 MÁS BAJA

Especie	Precisión	Recall	F1-score	Support
Anhinga anhinga_Anhinga	0.00	0.00	0.00	43
Anthrocothorax nigricollis_Black-throated Mango	0.00	0.00	0.00	19
Nomonyx dominicus_Masked Duck	0.00	0.00	0.00	14
Panyptila cayennensis_Lesser Swallow-tailed Swift	0.00	0.00	0.00	3
Porphyrio martinica_Purple Gallinule	0.00	0.00	0.00	75

En total, se seleccionaron 95 especies de aves y la categoría de voces humanas que cumplen con este umbral de F1-score. Estas especies representan un subconjunto de las que se encuentran en el corpus de sonidos de aves y son las que el modelo puede identificar con una precisión y recall razonables. Esta selección permite centrar los esfuerzos en la mejora de la identificación de estas especies, al tiempo que se reconoce la necesidad de futuros ajustes y optimizaciones para abordar con mayor eficacia las especies con F1-scores más bajos.

B. Resultados de la CNN 96 Categorías

Los resultados de la evaluación de la red convolucional entrenada muestran un progreso positivo a medida que se avanza a través de las épocas de entrenamiento. Inicialmente, en la primera época, tanto la pérdida como la precisión eran bajas, lo que es esperado en las primeras etapas de entrenamiento. A medida que el modelo se ajusta a los datos, se observa una mejora significativa en la pérdida y la precisión en las siguientes épocas. Esta tendencia alentadora sugiere que el modelo aprende y se adapta mejor a este conjunto de datos más reducido de sonidos de aves. Sin embargo, a medida que se avanza en el entrenamiento, se aprecia una disminución en la mejora de la precisión en el conjunto de validación, lo que indica que el modelo podría estar comenzando a sobreajustarse a los datos de entrenamiento. Esto evidencia que aún persiste un desequilibrio entre las clases en donde algunas especies son menos comunes.

En cuanto a las métricas finales, en la última época, la función de pérdida en el conjunto de prueba es de 2.44, y la precisión es del 45.75%. Las obtenidas para los conjuntos de entrenamiento y validación fueron de 1.90, 53.01% y 2.40, 47.91% respectivamente. Estas métricas reflejan la capacidad del modelo para clasificar los sonidos de aves en el conjunto de prueba. Aunque se ha logrado un progreso considerable en el rendimiento, aún hay margen para la mejora, especialmente en términos de precisión. A partir de estos resultados, se pueden tomar decisiones sobre cómo optimizar y ajustar el modelo en futuras iteraciones del entrenamiento. La Fig. 4 y la Fig. 5 ilustran el progreso del entrenamiento de la red.

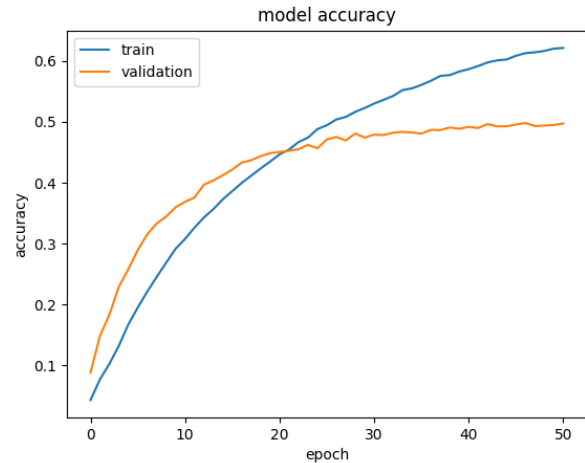


Fig. 4. CNN 96 categorías: Precisión para los conjuntos de entrenamiento y validación.

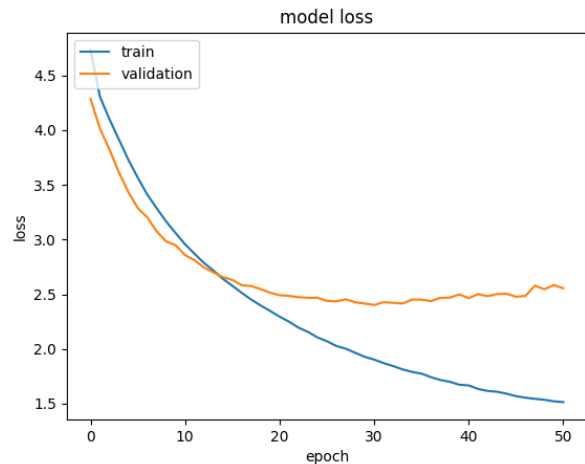


Fig. 5. CNN 96 categorías: Función de pérdida para los conjuntos de entrenamiento y validación.

C. Resultados de la Red Transformer 96 Categorías

A lo largo de las 60 épocas de entrenamiento, con la red Transformer se observa una tendencia general de mejora en la precisión del modelo en el conjunto de entrenamiento. La red comienza con una precisión bastante baja del 11.19% en la primera época y finaliza con una precisión del 65.07% en la última época. Esto sugiere que el modelo está aprendiendo y ajustándose mejor a medida que avanza el entrenamiento. En este mismo sentido, la función de pérdida en el conjunto de entrenamiento disminuye con el tiempo, lo que indica que el modelo se está volviendo más eficiente para minimizar la diferencia entre las predicciones y los valores reales.

Con respecto al conjunto de prueba, su rendimiento es esencial para evaluar su capacidad de generalización. En la última época, el modelo tiene una precisión del 60.89%, lo que sugiere que el modelo tiene un mejor desempeño al obtenido con la red convolucional. Esto también indica que el modelo no está sobreajustando (overfitting) gravemente los

datos de entrenamiento, ya que la precisión en el conjunto de prueba es razonable. A lo largo del entrenamiento, el modelo muestra signos de convergencia. La disminución de la función de pérdida y el aumento de la precisión en el conjunto de prueba son señales de que el modelo sigue mejorando y ajustándose, aunque quizás con una tasa decreciente. Dado que la precisión en el conjunto de prueba no alcanza su punto máximo en la última época, podría haber espacio para mejorar aún más el modelo a través de un entrenamiento adicional o la modificación de la arquitectura de la red. También es importante considerar el equilibrio entre precisión y eficiencia computacional, ya que los entrenamientos más largos pueden ser costosos en términos de recursos.

Aunque la red no parece estar sobreajustando gravemente, es importante realizar un seguimiento cuidadoso de las métricas de rendimiento a medida que se extiende el entrenamiento, ya que podría producirse sobreajuste si el modelo continúa entrenándose durante un período muy largo. Si el objetivo es mejorar aún más la precisión o la eficiencia, se pueden explorar diferentes técnicas, como el ajuste de hiperparámetros, el uso de arquitecturas de modelos más complejas o la recopilación de datos adicionales si es posible. La Fig. 6 y la Fig. 7 ilustran el progreso del entrenamiento de la red.

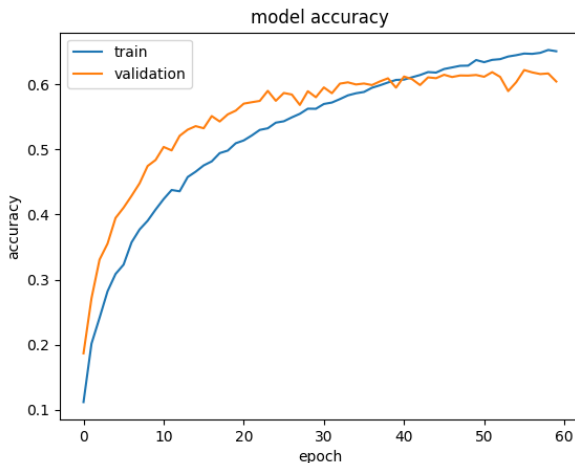


Fig. 6. Red Transformer 96 categorías: Precisión para los conjuntos de entrenamiento y validación.

D. Resultados Matemáticos

Los resultados obtenidos en la clasificación de sonidos de aves se expresan a través de la precisión del modelo basado en la arquitectura de Transformers, alcanzando un valor de $P_{\text{transformer}} = \frac{C_{\text{transformer}}}{T_{\text{transformer}}} = 65\%$, donde $C_{\text{transformer}}$ representa el número de clasificaciones correctas y $T_{\text{transformer}}$ el total de clasificaciones realizadas. La matriz de confusión para este modelo se puede representar como:

$$\text{Matriz de Confusión} = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix}$$

donde TP son los verdaderos positivos, TN los verdaderos negativos, FP los falsos positivos y FN los falsos negativos.

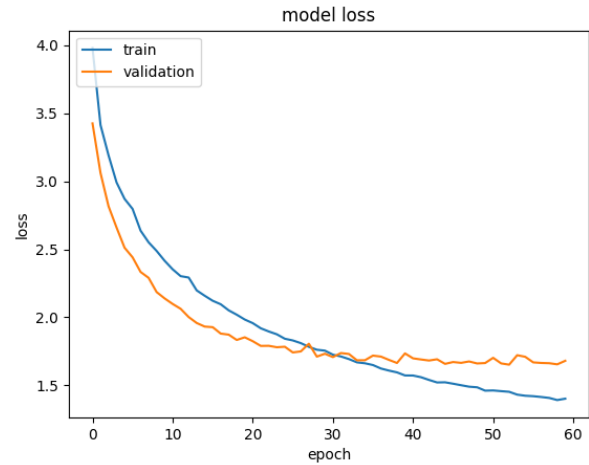


Fig. 7. Red Transformer 96 categorías: Función de pérdida para los conjuntos de entrenamiento y validación.

El error del modelo se calcula como $E_{\text{transformer}} = 1 - P_{\text{transformer}} = 35\%$, indicando que, a pesar de un rendimiento destacado, existe un margen de mejora en la identificación de los sonidos de aves. Estos resultados proporcionan una base para futuras investigaciones, sugiriendo que la inclusión de una mayor variedad de categorías de sonidos en el corpus puede mejorar aún más la precisión de los modelos de clasificación.

VI. PRUEBA DE CONCEPTO

La *Sicalis flaveola*, comúnmente conocida como Saffron Finch, Fig. 8, es un pequeño pájaro amarillo ampliamente presente en América del Sur. En el contexto de nuestro trabajo, se consideró el trabajo de [18], en el cual se evaluó el impacto del ruido vehicular sobre el canto y comportamiento de esta especie en la zona norte de la ciudad de Armenia, Quindío. El objetivo principal de este análisis fue el de determinar cómo las variaciones en los niveles de ruido influyen en la comunicación acústica y las actividades diarias de *Sicalis flaveola*. Los resultados de dicho estudio proporcionan un marco relevante para entender las adaptaciones y posibles vulnerabilidades de estas aves en entornos urbanos ruidosos.

La *Sicalis flaveola* es una especie frecuentemente avistada en áreas urbanas del departamento del Quindío, adaptándose notablemente bien a estos entornos. Su presencia es particularmente significativa en zonas verdes y parques, donde su canto característico añade un componente sonoro distintivo al paisaje urbano. Esta adaptabilidad convierte a la *Sicalis flaveola* en un excelente sujeto de estudio para investigar los efectos del ruido antropogénico en aves urbanas, y facilita la recopilación de datos necesarios para nuestro análisis de clasificación de sonidos. [18].

Durante una jornada de campo en el municipio de La Tebaida, Quindío, se llevó a cabo la grabación de los cantos de la *Sicalis flaveola*. Esta actividad se realizó en diferentes momentos del día para capturar una muestra representativa de las vocalizaciones de la especie en su hábitat natural. Las grabaciones obtenidas fueron posteriormente procesadas



Fig. 8. Saffron Finch (*Sicalis flaveola*).

para su análisis, proporcionando una base para evaluar el desempeño de la red entrenada.

De las grabaciones realizadas, se seleccionaron 36 audios que fueron considerados de alta calidad y representativos del canto de la *Sicalis flaveola*. A partir de estos audios, se extrajeron 108 espectrogramas. El ViT Transformer entrenado clasificó correctamente 71 de los 108 espectrogramas analizados, alcanzando una precisión del 65.74%. Este valor se aproxima al obtenido con el conjunto de prueba empleado para evaluar el modelo, correspondiente al 63.40%. Ver Tabla IV.

TABLA IV
ESPECTROGRAMAS POR ESPECIE DE AVE

Especie de ave	Espectrogramas
<i>Buteo nitidus</i> _Gray-lined Hawk	2
<i>Claravis pretiosa</i> _Blue Ground Dove	10
<i>Columbina talpacoti</i> _Ruddy Ground Dove	1
Black-bellied Whistling-Duck	6
<i>Falco sparverius</i> _American Kestrel	9
<i>Gampsonyx swainsonii</i> _Pearl Kite	1
<i>Legatus leucophaeus</i> _Piratic Flycatcher	1
<i>Sicalis flaveola</i> _Saffron Finch	71
<i>Tyrannus savana</i> _Fork-tailed Flycatcher	5
Voces humanas	1
<i>Xenops rutilans</i> _Streaked Xenops	1

En el análisis de los espectrogramas clasificados por la red Transformer y presentados en la Tabla IV, se observa una frecuencia apreciable en la identificación para las especies *Falco sparverius* (American Kestrel) y *Claravis pretiosa* (Blue Ground Dove). Esta frecuencia puede atribuirse a ciertas semejanzas en sus cantos, que podrían incluir patrones acústicos comunes o características espectrales similares que la red neuronal reconoce con mayor facilidad. Estas especies, aunque pertenecen a diferentes familias y tienen distintos comportamientos y hábitats, pueden compartir frecuencias de llamada similares, modulaciones de tono o duraciones de las notas que son destacables en sus cantos.

Específicamente, la *Claravis pretiosa* y la *Sicalis flaveola* suelen producir cantos melódicos con repetición de notas en intervalos regulares, lo que podría facilitar su clasificación. Por otro lado, el *Falco sparverius*, aunque es un ave de presa,

emite llamadas que pueden tener una estructura rítmica y tonal comparable en ciertos contextos, especialmente en llamadas de alerta o comunicación intraespecífica. Estas similitudes acústicas subrayan la importancia de considerar tanto las características físicas de los cantos como los contextos ecológicos en los que se producen para entender mejor los resultados de clasificación de modelos basados en inteligencia artificial.

VII. CONCLUSIONES

El departamento del Quindío en Colombia es una región rica en avifauna autóctona, y esta diversidad biológica contribuye significativamente a la biodiversidad del país. La conservación y el estudio de estas especies son esenciales para la preservación de esta riqueza natural. En este contexto, se ha creado un corpus de sonidos que incluye audios de aproximadamente el 30% de las especies de aves identificadas en el Quindío, junto con otros sonidos como voces humanas, silencios y ruidos. Este corpus es una herramienta valiosa para el estudio y la conservación de la avifauna regional.

Para analizar y clasificar los sonidos de aves en este corpus, se empleó un enfoque de “extracción de características de ventana deslizante”. Además, se utilizaron redes neuronales para evaluar el corpus. En particular, se entrenaron tres redes neuronales, incluida una basada en la arquitectura de Transformers. Los resultados mostraron niveles de precisión en la clasificación de sonidos de aves, con un rendimiento destacado de la red basada en Transformers, que obtuvo una precisión del 65%. Esto sugiere que la arquitectura de Transformers es especialmente efectiva en esta tarea y puede ser una opción prometedora para futuras investigaciones.

Estos resultados ofrecen una línea base para futuras investigaciones en el campo de la clasificación de sonidos de aves. No obstante, se sugiere ampliar el número y su volumen de categorías de sonidos a incluir en el conjunto de datos. Esto implica incorporar una variedad de sonidos naturales y artificiales, ya que estos son factores importantes en el entorno sonoro en el que se encuentran las aves. Al considerar una gama más amplia de ruidos ambientales, los modelos propuestos serán capaces de distinguir y filtrar de manera más precisa los cantos de las aves en contextos reales. Al mejorar la capacidad de identificar y clasificar sonidos de aves autóctonas, se está contribuyendo a la conservación de la avifauna regional y, por extensión, a la preservación de la biodiversidad en el Quindío y en Colombia en su conjunto.

VIII. REPRODUCIBILIDAD

Esperamos que este trabajo ayude a impulsar investigaciones futuras sobre aprendizaje automático para la clasificación de especies de aves a partir del monitoreo acústico pasivo. Por esta razón, el código fuente utilizado en los experimentos presentados en este documento se encuentra disponible en el siguiente repositorio:⁴. Los conjuntos de datos están disponibles según se indicó en la Sección IV.

⁴<https://github.com/helbert-novoa/QuindiBIRD>

REFERENCIAS

- [1] S. Kahl, Connor M. Wood, Maximilian Eibl, & Holger Klinck. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61, 101236. doi: 10.1016/j.ecoinf.2021.101236.
- [2] INECOL Instituto de Ecología, AC. (2021). Aves, señales acústicas y paisajes sonoros. <https://elportal.mx/salud-y-ciencia/inacol-es-ciencia/08/03/2021/aves-senales-acusticas-y-paisajes-sonoros/>.
- [3] Instituto de Investigación de Recursos Biológicos Alexander von Humboldt. (2020). Conjunto de datos de monitoreo acústico pasivo en la Reserva Natural Los Yátaros de Gachantivá, Boyacá, Colombia. 12447 eventos. <https://doi.org/10.15472/wt0elp>.
- [4] J. Salamon, J. P. Bello, A. Farnsworth, & S. Kelling. (2017). Fusing shallow and deep learning for bioacoustic bird species classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 141-145). doi: 10.1109/ICASSP.2017.7952134.
- [5] B. Sanchez, F.J., Hossain, M.R., English, N.B., et al. (2021). Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture. *Scientific Reports*, 11, 15733. <https://doi.org/10.1038/s41598-021-95076-6>.
- [6] Kiskin, I., Zilli, D., Li, Y., et al. (2020). Bioacoustic detection with wavelet-conditioned convolutional neural networks. *Neural Computing and Applications*, 32, 915–927. <https://doi.org/10.1007/s00521-018-3626-7>.
- [7] Potamitis, I., Ganchev, T., Kontodimas, D. (2009). On Automatic Bioacoustic Detection of Pests: The Cases of *Rhynchophorus ferrugineus* and *Sitophilus oryzae*. *Journal of Economic Entomology*, 102(4), 1681–1690. <https://doi.org/10.1603/029.102.0436>.
- [8] Zhang, S., Gao, Y., Cai, J., Yang, H., Zhao, Q., & Pan, F. (2023). A Novel Bird Sound Recognition Method Based on Multifeature Fusion and a Transformer Encoder. *Sensors*, 23(19), 8099. <https://doi.org/10.3390/s23198099>.
- [9] Nagesh, C. K., & Purushothama, A. (2022). The Birds Need Attention Too: Analyzing usage of Self Attention in identifying bird calls in soundscapes. [Preprint]. <https://doi.org/10.48550/arXiv.2211.07722>.
- [10] Puget, J. (2021). STFT Transformers for Bird Song Recognition. *Conference and Labs of the Evaluation Forum*.
- [11] Gómez, W.E., Isaza, C.V., & Daza, J.M. (2018). Identifying disturbed habitats: A new method from acoustic indices. *Ecological Informatics*, 45, 16-25. <https://doi.org/10.1016/j.ecoinf.2018.03.001>.
- [12] M. Sugai, L. S., Freire Silva, T. S., Ribeiro, J. W. Jr., & Llusia, D. (2019). Terrestrial Passive Acoustic Monitoring: Review and Perspectives. *BioScience*, 69(1), 15–25. <https://doi.org/10.1093/biosci/biy147>.
- [13] Kalan, A.K., Mundry, R., Wagner, O.J., Heinicke, S., Boesch, C., & Kühl, H.S. (2015). Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecological Indicators*, 54, 217-226.
- [14] Browning, E., Gibb, R., Glover-Kapfer, P., & Jones, K.E. (2017). Passive acoustic monitoring in ecology and conservation. Woking, UK: WWF-UK. (WWF Conservation Technology Series 1(2)).
- [15] Bedoya, C., Isaza, C.V., Daza, J.M., & López, J.D. (2014). Automatic recognition of anuran species based on syllable identification. *Ecological Informatics*, 24, 200-209. <https://doi.org/10.1016/j.ecoinf.2014.08.009>.
- [16] Páez, J. P. C. (2022, 21 junio). Escuche aquí los paisajes acústicos del Golfo de Tribugá. *Revista Pesquisa Javeriana*. <https://www.javeriana.edu.co/pesquisa/el-golfo-de-tribuga-colombia-sonidos-paisajesacuaticos/>.
- [17] Bakker, K. (2022). Science is making it possible to ‘hear’ nature. It does more talking than we knew. *The Guardian*. <https://www.theguardian.com/commentsfree/2022/nov/30/science-hear-nature-digital-bioacoustics>.
- [18] R. Cabezas, M., & Riveros Bonilla, A. (2019). Análisis del canto y comportamiento de *Sicalis flaveola* (Aves: Thraupidae) en un gradiente de ruido vehicular en la ciudad de Armenia, Quindío. Disponible en: <https://bdigital.uniquindio.edu.co/handle/001/6055>.



Helver Novoa is a professional in statistics, holding a master's degree in engineering with an emphasis on software, and another master's degree in business intelligence. His career has primarily developed in the governmental sector, where he has worked as a statistician, database analyst, and data scientist. These experiences have allowed him to acquire a deep knowledge in handling and analyzing large volumes of information, optimizing decision-making processes. In addition to his work in the public sector, Helver also serves as a university lecturer.

His academic and professional focus centers on the integration of advanced statistical techniques with software tools, contributing to the development of innovative solutions in business intelligence.



Alexandra Rengifo Román is a student in her final semester of Systems and Computer Engineering from the University of Quindío. She is certified as a Google Cloud Professional Cloud Architect and works in that area. Her professional interests include software development, cloud architecture, data analysis, and research. In 2022, she represented Quindío in the event “Closing Gender Gaps in CTel,” an initiative of the Ministry of Science, Technology and Innovation in alliance with the Organization of Ibero-American States for Education, Science and Culture (OEI). Additionally, she is certified in STEAM/ICT Skills and in project formulation with an emphasis on research.



Emilio Granell obtained his B.Sc. degree in Telecommunications Engineering with the speciality in Sound and Image in 2006, his M.Sc. in Artificial Intelligence, Pattern Recognition, and Digital Image in 2011, and his Ph.D in Computer Science in 2017, all from Universitat Politècnica de València (UPV). He worked in a telecommunications consulting company in France from 2006 to 2007. Dr. Emilio Granell pertains to the Pattern Recognition and Human Language Technology (PRHLT) research center, where he develops his research on the topics dialogue systems, and interactive and multimodal systems. From 2010 he has participated in several research projects related with artificial intelligence, speech and handwritten text recognition, and smart cities. Currently, he is working in Sciling S.L. as a senior researcher.



Fáber D. Giraldo is a System and Computer Engineer from the University of Quindío, Colombia (with a grant from the Ministry of Education of Colombia). He has a Ms.Eng. degree with emphasis on Informatics from EAFIT University, Colombia (with a grant from EAFIT University). He holds a Ph.D. in Informatics from the Universidad Politècnica de València, Spain (with a grant from the National administrative department of Science, Technology and Innovation of Colombia – COLCIENCIAS, now Ministry of Science, Technology and Innovation). He

is an associate professor at the Faculty of Engineering of the University of Quindío, and also, he is Vice Rector of Research at the University of Quindío. His research interests — include — software — engineering, model-based engineering, software quality, model-based engineering quality, software architecture, enterprise architecture and HCI. ORCID: <http://orcid.org/0000-0002-6111-3055>.