

Escuela Superior Politécnica del Litoral

Facultad de Ingeniería en Electricidad y Computación

Reconocimiento de audios de aves del Bosque Protector La Prosperina utilizando técnicas de aprendizaje profundo.

TECH-337

Proyecto Integrador

Previo la obtención del Título de:

Ingeniero/a en Ciencias de la Computación

Presentado por:

Tyrone Eduardo Rodríguez Motato

Adriana Brigitte Guilindro Garcés

Guayaquil - Ecuador

2023 - 2024

Dedicatoria

Con profundo cariño, dedico este proyecto a la memoria de mi abuelita Flor Amalia Espinoza de los Monteros, quien siempre creyó en mis capacidades, estoy aquí gracias a su eterna fe y amor. A mis padres, Cinthya Motato y Eduardo Rodriguez, les dedico este logro, agradeciendo su apoyo incondicional que ha sido la base de mi crecimiento. A mi querida hermana Amaia Rodriguez, quien siempre ha estado a mi lado. A la que considero mi segunda madre mi Tita Cecibel, por su cuidado y sabios consejos. A Emily Pérez, pilar fundamental en mis dos últimos años, donde su cariño impulsa mi superación diaria, inspirándome a creer que yo, puedo con todo.

Tyrone Rodriguez

Dedicatoria

El presente proyecto lo dedico a mi madre Mónica Garcés, a mi padre Julio César Guilindro, a mi hermana Priscilla Guilindro y a mis abuelos Fanny Cruz y Juan Garcés que me ha apoyado a lo largo de toda mi vida personal y académica, a Blanquita, Molly, Estela y Yuumi, mis cuatro gatas que, con sus ocurrencias, volvían los momentos más estresantes superables, a la banda BTS que con sus canciones me han inspirado para seguir adelante y cumplir mis metas.

Adriana Guilindro

Agradecimientos

Quiero expresar mi sincero agradecimiento al Dr. Miguel Realpe, nuestro tutor, por su apoyo invaluable. Su orientación fue esencial para alcanzar nuestros objetivos y más. Agradezco al Centro de Investigación CIDIS por facilitar la postulación del próximo artículo científico.

Al Dr. Paolo Piedrahita por sus valiosos aportes a la conservación de aves.

Mi agradecimiento se extiende a mi familia por su afecto y bendiciones. A mi grupo de amigos, con quienes compartí desde primer semestre, y con los que recién conocí en los últimos años de mi carrera, aportándome no sólo conocimiento, sino momentos extremadamente felices.

Tyrone Rodriguez

Agradecimientos

Quiero dar mis más sinceros agradecimientos al Dr. Miguel Realpe como tutor de este proyecto por su dedicación a resolver, revisar y dar retroalimentación en cada una de las etapas para llegar a este resultado.

Al Dr. Paolo Piedrahita por sus aportes para comprender el estado de la importante labor de la conservación de las aves.

Al Msc, Rodrigo Alexander Saraguro Bravo, por su labor como consejero académico durante gran parte de mi etapa estudiantil desde el inicio de la pandemia.

Agradezco profundamente el apoyo incondicional que me ha brindado mi familia y a mi compañero de integradora Tyrone Rodríguez por acompañarme en este trabajo.

Adriana Guilindro

Declaración expresa

Nosotros, Tyrone Eduardo Rodríguez Motato y Adriana Brigitte Guilindro Garcés, acordamos y reconocemos que:

La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores. La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por mí/nosotros durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que me/nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de mi/nuestra innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique los autores que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 16 de febrero del 2024



Tyrone Eduardo Rodríguez Motato



Adriana Brigitte Guilindro Garcés

Evaluadores

.....
Ph.D. Lucia Marisol Villacres Falconi

Profesor de la materia

.....
Ph.D. Miguel Andres Realpe Robalino

Profesor tutor

Resumen

Este proyecto se centra en desarrollar un sistema innovador para la conservación y comprensión de la biodiversidad aviar en el ecosistema del Bosque Protector "La Prosperina" en Ecuador. Implementa técnicas avanzadas de aprendizaje profundo, específicamente utilizando un enfoque de auto decodificadores variables (VAEs) y Agrupamiento. Se aplican técnicas de preprocesamiento, como la separación de señales de ruido, y se diseñan modelos basados en auto decodificadores variables variacionales para extraer características significativas de los espectrogramas de Mel. El modelo entrenado demuestra habilidades excepcionales para identificar especies de aves amenazadas, con resultados positivos en las métricas de evaluación y altas tasas de precisión. Además de beneficiar a los conservacionistas al proporcionar herramientas efectivas y económicamente viables para el monitoreo de especies y comportamientos aviares, el proyecto tiene aplicaciones más amplias, incluida la conciencia ambiental y educación.

Palabras claves: Reconocimiento de aves, Aprendizaje profundo, Bosque Protector "La Prosperina", VAEs, Clustering.

Abstract

This project focuses on developing an innovative system for the conservation and understanding of avian biodiversity in the ecosystem of the "La Prosperina" Protective Forest in Ecuador. It implements advanced deep learning techniques, specifically utilizing a variational autoencoder and clustering approach. Preprocessing techniques, such as noise signal separation, are applied, and variational autoencoder models are designed to extract meaningful features from Mel spectrograms. The trained model demonstrates exceptional abilities to identify threatened bird species, with positive results in the evaluation metrics and high accuracy rates. In addition to benefiting conservationists by providing effective and economically viable tools for monitoring bird species and behaviors, the project has broader applications, including environmental awareness and education.

Keywords: *Birds Species identification, Deep Learning, La Prosperina Protective Forest, VAEs, Clustering*

Índice general

Evaluadores	VII
Resumen	VIII
Abstract.....	IX
Índice general	X
Índice de Figuras.....	XII
Índice de Tablas	XII
Índice de Ecuaciones	XII
Abreviaturas	XIII
Simbología	XIV
Capítulo 1.....	15
1.1 Introducción.....	16
1.2 Descripción del problema.....	16
1.3 Justificación del problema.....	18
1.4 Objetivos	19
1.4.1 Objetivo general	19
1.4.2 Objetivos específicos.....	19
1.5 Marco teórico.....	19
1.5.1 Estado del Arte	19
1.5.2 Auto codificadores Variacionales	21
1.5.3 Espectrogramas.....	22
1.5.4 Preprocesamiento de audio	23
1.5.5 Estudio acústico.....	24
1.5.6 Proyectos similares	25
Capítulo 2.....	26
2.1 Metodología.....	27
2.1.1 Análisis y definición del problema	27
2.1.2 Requerimientos.....	28

2.1.3	Usuarios y beneficiarios de la solución	29
2.1.4	Diseño de la solución	30
2.1.5	Base de datos.....	31
2.1.6	Aumento de data.....	31
2.1.7	Arquitectura	33
2.1.8	Detección de anomalías	34
2.1.9	Librerías	34
2.1.10	Interfaz web.....	35
2.1.11	Diagramas	36
Capítulo 3.....		40
3.1	Plan de evaluación.....	41
3.2	Resultados y análisis.....	42
3.3	Interfaz Web	45
3.4	Análisis de Costos y viabilidad.....	47
3.5	Aporte Académico-Científico.....	48
3.6	Publicación de artículo científico.....	48
Capítulo 4.....		49
4.1	Conclusiones y recomendaciones.....	50
4.1.1	Conclusiones	50
4.1.2	Recomendaciones	50
Referencias.....		52

Índice de Figuras

Figura 1.....	20
Figura 2.....	21
Figura 3.....	23
Figura 4.....	24
Figura 5.....	28
Figura 6.....	30
Figura 7.....	37
Figura 8.....	38
Figura 9.....	39
Figura 10.....	41
Figura 11.....	42
Figura 12.....	43
Figura 13.....	44
Figura 14.....	45
Figura 15.....	46
Figura 16.....	46

Índice de Tablas

Tabla 1.....	17
Tabla 2.....	47

Índice de Ecuaciones

Ecuación 1.....	22
-----------------	----

Abreviaturas

ESPOL Escuela Superior Politécnica del Litoral

IUCN Unión Internacional para la Conservación de la Naturaleza

DL Aprendizaje Profundo

EN En peligro

VU Vulnerable

NT Casi Amenazada

ARU Estaciones de audio autónomas

VAE Variational Autoencoder

RMS Raíz Cuadrada Media

MRR Mean Reciprocal Rank

DP Procesamiento digital

STFT Transformada de Fourier a Corto Plazo

MFCC Coeficientes cepstrales de frecuencia Mel

GPU Unidad de Procesamiento Gráfico

ICICT International Conference on Information and Computer Technologies

Simbología

f	Frecuencia
m	Escala de Mel

CAPÍTULO 1

1.1 Introducción

Las aves son unos de los agentes más importantes para el equilibrio de los ecosistemas y de nuestro sistema económico social actual, debido a que son polinizadores naturales, lo que ayuda a la reproducción de especies de flora, controlan plagas que pueden afectar cultivos y también sirven como indicadores de contaminación. Por lo que, tener un seguimiento de las especies que se encuentren en mayor o menor medida de amenaza de extinción, debido a diversas razones, se considera el componente central en la disciplina de la ornitología [1], teniendo en cuenta que es posible que no existan planes efectivos para su conservación.

1.2 Descripción del problema

La conservación de especies de aves se ha convertido en una de las principales preocupaciones en Ecuador, ya que es uno de los cinco países con mayor diversidad de aves del mundo, con 1.699 endémicas, migratorias y residentes, entre otras. Según las categorías de riesgo de extinción definidas por la Unión Internacional para la Conservación de la Naturaleza (IUCN) [2], entre el 10% y el 20% de estas especies están en peligro de extinción. Sin embargo, algunas especies de aves, aunque clasificadas por la IUCN como "preocupación menor", reclasificadas como "En Peligro" según la Lista Roja de Aves de Ecuador [3], teniendo un total 312 especies calificadas como bajo algún grado de amenaza [4].

Este proyecto tiene como objetivo identificar especies de aves en peligro de extinción a través de grabaciones de audio ambiental, automatizando la detección de especies mediante el uso de métodos de Aprendizaje Profundo (DL). Considerando que los métodos tradicionales empleados en la monitorización de aves requieren que los conservacionistas calculen las poblaciones e identifiquen las aves escuchando grabaciones de audio, un proceso caracterizado por su alto costo en mano de obra [7, 8, 9]. Al mismo tiempo, las grabaciones ambientales pueden ser excepcionalmente extensas y humanamente inviables de escuchar en su totalidad. Además, una parte significativa de estas grabaciones contiene sonidos no interesantes, como viento, lluvia, otros animales, etc.

Para seleccionar las especies de interés, se consideraron trece especies actualmente en peligro en el Bosque de Protección "La Prosperina". Además, debe haber un número considerable de sus vocalizaciones disponibles en bases de datos como "xeno-canto.org" [10].

Tabla 1

Especies amenazadas en el Bosque Protector "La Prosperina"

Nombre común	Nombre científico	Familia	Estado de conservación
Gavilán Dorsigrís	<i>Pseudastur occidentalis</i>	<i>Accipitridae</i>	EN
Amazona Frentirroja	<i>Amazona autumnalis</i>	<i>Psittacidae</i>	EN
Anambé Pizarra	<i>Pachyramphus spodiurus</i>	<i>Cotingidae</i>	EN
Chachalaca Cabecirrufa	<i>Ortalis erythroptera</i>	<i>Cracidae</i>	VU
Perico Cachetigrís	<i>Brotogeris pyrrhoptera</i>	<i>Psittacidae</i>	VU
Eufonia Coroninaranja	<i>Euphonia saturata</i>	<i>Passeriformes</i>	VU
Urraca Coliblanca	<i>Cyanocorax mystacalis</i>	<i>Corvidae</i>	VU
Carpintero Guayaquileño	<i>Campephilus gayaquilensis</i>	<i>Picidae</i>	VU
Trogón Ecuatoriano	<i>Trogon mesurus</i>	<i>Trogonidae</i>	VU
Picolete Ecuatoriano	<i>Picumnus sclateri</i>	<i>Picidae</i>	NT
Perico Caretirrojo	<i>Psittacara erythrogenys</i>	<i>Psittacidae</i>	NT

Tinamú Cejipalido	<i>Crypturellus transfasciatus</i>	<i>Sthruotioniformes</i>	NT
Mosquero Pechigris	<i>Lathrotriccus griseipectus</i>	<i>Tyrannidae</i>	NT

Nota: El estado de conservación utiliza la nomenclatura de la Unión Mundial para la Conservación de la Naturaleza: “En Peligro” (EN), “Vulnerable” (VU), “Casi Amenazada” (NT).

1.3 Justificación del problema

La implementación de un proyecto de reconocimiento del canto de las aves tendría una serie de beneficios e impactos para temas de preservación a nivel mundial. Puesto que, ayudaría a los conservacionistas a estudiar las poblaciones de aves y a seguir los cambios en su comportamiento a lo largo del tiempo, esta información podría utilizarse para desarrollar estrategias de conservación más eficaces. A su vez, ayudaría a desarrollar nuevas tecnologías para vigilar y controlar las plagas y enfermedades que propagan las aves, con ello se podrían reducir los daños económicos y medioambientales causados por estas plagas y enfermedades [11].

Por último, se podrían crear nuevos productos educativos y de entretenimiento, como aplicaciones de identificación de aves y documentales sobre la naturaleza, cada producto permitiría que la gente aprenda más sobre las aves y el mundo natural de nuestra universidad. La implementación de este proyecto beneficiará a un gran abanico de personas, como conservacionistas, científicos, estudiantes y el público.

1.4 Objetivos

1.4.1 Objetivo general

Reconocer especies de aves que habitan en el ecosistema de Bosque seco mediante la aplicación de técnicas inteligentes basadas en aprendizaje automático para el apoyo de planes de conservación en la ESPOL.

1.4.2 Objetivos específicos

1. Identificar especies de aves que habitan en el ecosistema de Bosque seco que se encuentren en algún grado de amenaza.
2. Recopilar el registro de las aves identificadas en base a sus cantos.
3. Diseñar la estructura de un algoritmo que permita el ingreso de un audio extenso y devuelva como resultado las especies de aves presentes en el mismo.
4. Desarrollar el algoritmo de reconocimiento de cantos de las especies de aves de interés a través de técnicas de aprendizaje profundo.
5. Evaluar el desempeño del algoritmo de reconocimiento para validar su utilidad en el apoyo de planes de conservación.

1.5 Marco teórico

1.5.1 Estado del Arte

La bioacústica es la ciencia que combina dos ramas científicas que son la acústica y la biología para estudiar los comportamientos comunicativos que existen entre los animales en un hábitat o ecosistema [12]. Mayormente, este campo de estudio implica la investigación desde cómo se produce un sonido por un emisor, cómo se propaga a través de un medio y como es recibido por el receptor. Sus aplicaciones prácticas están en gran medida relacionadas con la conservación y manejo de especies, además de estudios relacionados a la descripción de nuevas especies, entre otros.

Para extraer estos sonidos, se suelen instalar grabadoras o estaciones de audio autónomas (ARU) en lugares estratégicos para obtener la mayor información posible según qué especies se desean estudiar, por ejemplo: si viven cerca de la tierra como anfibios o mamíferos, las grabadoras estarán cerca del suelo

y sin afectar los hábitos de la fauna del sector. En ciertas ocasiones, ya sea por falta de recursos o por acciones voluntarias de conservacionistas, algunas grabaciones se realizan a través de teléfonos o grabadoras menos sofisticadas mientras se recorre la zona a pie [13].

Figura 1

Grabadora de audio autónomas instalada en un bosque



Nota: Cortesía de Marconi Campos / Sieve Analytics

A nivel internacional, se conoce el proyecto “Rainforest Connection”, una fundación sin fines de lucro que instala este tipo de grabadoras en diferentes bosques a nivel mundial para ayudar a planes para combatir la deforestación ilegal, el monitoreo de especies y localización de especies en peligro de extinción en áreas estratégicas, además de brindar la experiencia de escuchar las grabaciones a través de su página web y su aplicación móvil como parte de su visión de proteger los ecosistemas en 34 países diferentes [14].

A nivel de ESPOL, se cuenta con el proyecto “Creación y Operación de una red de sensores acústicos y ambientales”, liderado por Pedro Villegas y Gabriel Intriago, quiénes en 2018 instalaron diez cámaras trampa a lo largo del Bosque Protector “La Prosperina” con micrófonos para estudiar la fauna del sector, según los encargados, se realizan capturas de un minuto y medio por cada cinco minutos, además de otras detecciones como la humedad y la radiación solar [15].

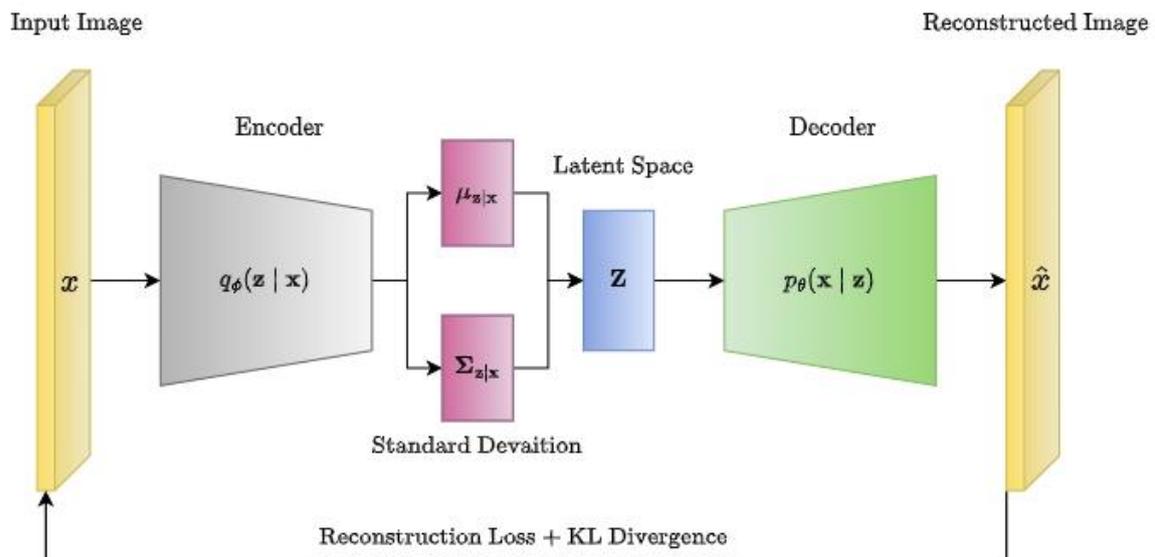
1.5.2 Auto codificadores Variacionales

Se le conoce como VAE o auto codificadores variacionales a una tipología de modelos gráficos probabilísticos que se basan en RN. Estos permiten construir inferencias que sean eficientes ante variables latentes discretas o continuas, distribución intratable del posterior para aquellos conjuntos de datos de un tamaño considerablemente grande [16].

Además, poseen la capacidad de la generación de nuevos datos que son similares a los de la distribución de origen mediante la codificación, en forma de variables latentes, de las características definitorias de la distribución de probabilidad que está representada por la muestra poblacional del conjunto original de datos. Por tanto, a partir de un modelo aprendido, pueden generar nuevas instancias que se aproximan a dicha distribución [17].

Figura 2

Arquitectura de un Variational Autoencoder (VAE)



Nota: Cortesía de The Medium

Una red codificadora (también llamada red de reconocimiento o encoder) transforma la distribución de origen a un espacio latente estocástico representado por una distribución más sencilla (ej: una distribución normal). A partir de la distribución latente se generan los datos de reconstrucción mediante el uso de una red decodificadora (o red de generación o decoder). Al finalizar el entrenamiento, si este fue

efectivo, será posible generar nuevas instancias mediante el decoder, transformando las muestras del espacio oculto al espacio de expresión. El encoder actúa como un modelo de reconocimiento, mientras que el decoder como uno de generación. El modelo de reconocimiento proporciona al generador una aproximación del posterior sobre el espacio latente, que es utilizado para optimizar los parámetros de ambas redes mediante estimación por máxima verosimilitud. Según la regla de Bayes, el modelo de reconocimiento aproxima a la inversa del modelo generativo. [18]

1.5.3 Espectrogramas

Los espectrogramas son representaciones visuales del contenido de frecuencia de un sonido a lo largo del tiempo. Convierten las señales de audio en imágenes bidimensionales, donde el eje horizontal representa el tiempo, el eje vertical representa las frecuencias, y el color o intensidad indica la amplitud de las frecuencias en cada momento [19], como se muestra en la Figura 3. Esto da una representación visual que pueden entender fácilmente los modelos de DP.

Además, la escala Mel tiene en cuenta la percepción no lineal de las frecuencias en el oído humano [20]. Esto hace que el espectrograma Mel sea más útil para capturar detalles relevantes en diferentes partes del espectro de frecuencia. Entre todas las conversiones de señales de sonido, los espectrogramas Mel destacan como la elección más efectiva en aplicaciones de Aprendizaje Profundo (DL) para lograr resultados óptimos [21]. Por lo tanto, son preferidos en muchas aplicaciones de procesamiento de señales de audio y reconocimiento de voz porque se asemejan más a cómo los humanos perciben el sonido.

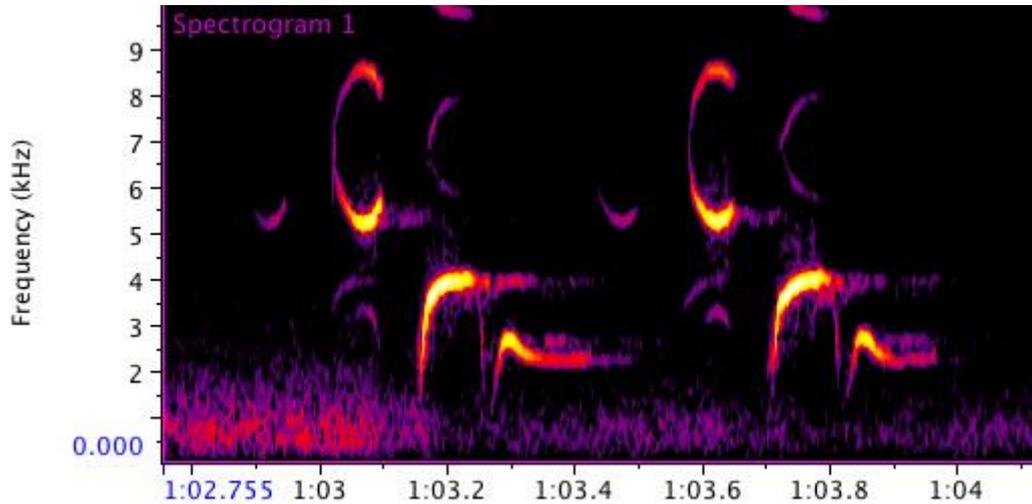
La conversión de la frecuencia (f) a la escala Mel (m) se da por:

$$m = 2595 \cdot \log \left(1 + \frac{F}{500} \right) \quad (1)$$

Por lo tanto, un espectrograma Mel es un espectrograma en el que las frecuencias se convierten a la escala Mel [22].

Figura 3

Canción de Oak Titmouse con una primera nota bitonal.



Nota: Espectrograma obtenido de [22].

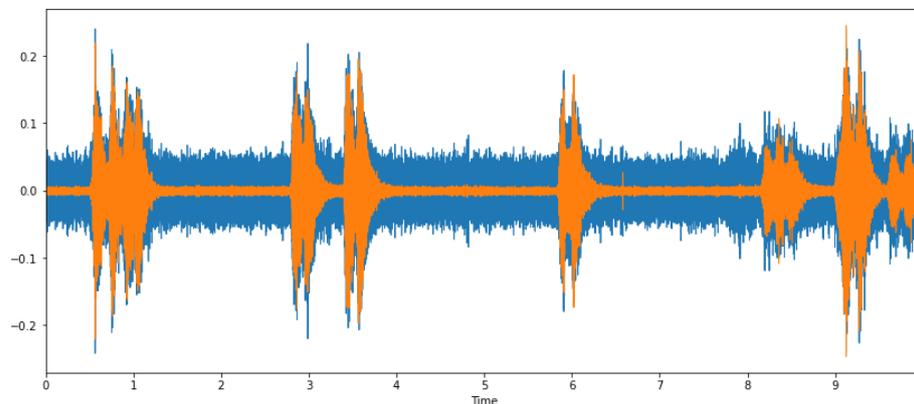
1.5.4 Preprocesamiento de audio

Para el preprocesamiento de datos, es esencial considerar tanto la reducción de ruido como la normalización, ya que son pasos fundamentales en la preparación de archivos de audio para el reconocimiento de cantos de aves. Este proceso implica manipular o transformar señales de sonido para eliminar ruidos no deseados y ajustar aspectos de tiempo y frecuencia mediante la conversión de señales digitales y analógicas. La reducción de ruido es el proceso de eliminar ruidos no deseados de una señal de audio, que pueden ser causados por diversos factores como personas, viento y otros sonidos ambientales, como se ilustra en la Figura 4. En el artículo [23], se habla de diversas técnicas de reducción del ruido que pueden utilizarse para el reconocimiento del canto de las aves. Algunas de las técnicas mencionadas son

- Promedio: técnica que consiste en promediar varias copias de la señal de audio ayudando a reducir el ruido que no está presente en todas las copias de la señal.
- Filtrado: técnica que consiste en filtrar frecuencias específicas que se sabe que están asociadas al ruido.
- Sustracción espectral: técnica que consiste en sustraer una estimación de ruido de la señal de audio.

Figura 4

Reducción de ruido usando la librería ‘Librosa’ en Python.



Nota: Cortesía de Silas Rudolf

Por otra parte, la normalización es el proceso de ajustar la amplitud de una señal de audio a un nivel estándar. Esto garantiza que las características extraídas de la señal sean comparables en diferentes grabaciones. De manera similar, en el artículo [24], se discuten los efectos de diferentes técnicas de normalización, como la Normalización de Pico y la Raíz Cuadrada Media (RMS), en relación con el rendimiento de los modelos de DP. El estudio concluye que la normalización puede mejorar la precisión de los modelos de aprendizaje automático para el reconocimiento de cantos de aves, especialmente en entornos ruidosos.

Este procedimiento es vital, pues se basa en manipular o transformar señales sonoras, eliminar el ruido no deseado y ajustar el tiempo y frecuencia transformando señales digitales y analógicas [25]. Con esto, la aplicación de Deep Learning en el análisis de audio ha experimentado un crecimiento significativo en los últimos años y ha abierto nuevas posibilidades en una amplia variedad de campos. Algunos ejemplos incluyen el reconocimiento de voz, etiquetado automático de música, clasificación, huellas dactilares de audio, segmentación y generación de audio [25].

1.5.5 Estudio acústico

En el estudio acústico, el método de investigación y seguimiento ecológico utilizado, especialmente para monitorear poblaciones y ecosistemas, y para registrar el comportamiento animal [26]. En este

contexto, las ARU desempeñan un papel fundamental, ya que pueden grabar sonidos en varios entornos. Los conservacionistas adoptan cada vez más estas herramientas para realizar estudios acústicos de aves y otros organismos, lo que les permite recopilar información valiosa para conservar y comprender la vida silvestre [26].

1.5.6 Proyectos similares

El estudio [27] presenta un sistema de identificación de aves basado en audio que utiliza un codificador de transformación compuesto, operaciones residuales y normalización para procesar las características del sonido de las aves y capturar relaciones posicionales espaciales. Se utilizó un conjunto de datos de grabaciones de audio de aves que abarca diversas propiedades, como diferentes frecuencias de muestreo y calidad de sonido, para el entrenamiento. Se aplicaron varias técnicas de preparación y aumento de datos, como la reducción de ruido de fondo y la manipulación de la duración del audio. Además, el modelo se evaluó en dos conjuntos de tareas: la identificación de aves en grabaciones de audio mono y en paisajes sonoros. Los resultados muestran un puntaje MRR (Mean Reciprocal Rank) del 82.7%. Además, la inclusión de predicciones para especies menos probables y la selección de instantáneas de modelos previos contribuyeron significativamente a mejorar el rendimiento.

Por otro lado, [28] presentó un enfoque de Aprendizaje Profundo llamado BirdNET para la identificación de aves, basado en datos de monitoreo acústico pasivo. Se basó en la extracción y procesamiento de espectrogramas de audio, con parámetros ajustados para una mejor representación de las características acústicas de las aves. El modelo se construyó sobre una red residual (ResNet) de 157 capas y se entrenó en alrededor de 1.5 millones de espectrogramas de entrenamiento. Los resultados demostraron que BirdNET logró una fuerte correlación con las observaciones humanas en la identificación de especies de aves, tanto en grabaciones enfocadas como en grabaciones de paisajes sonoros.

CAPÍTULO 2

2.1 Metodología

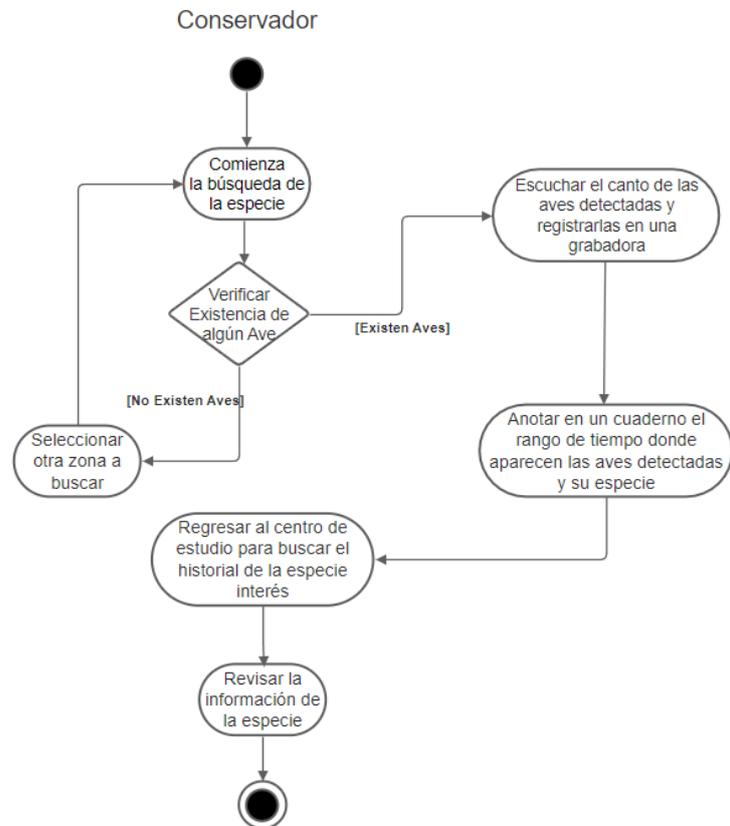
En este capítulo se presenta la metodología para desarrollar la solución propuesta, que abarca el análisis y definición del problema, los requerimientos funcionales y no funcionales, reconocimiento de los usuarios beneficiarios, diseño de la solución con datos de audios de aves, procesamiento de datos y arquitectura del modelo de inteligencia artificial, además de especificar las librerías y detallar la implementación. Todo esto se presenta a través de diagramas UML en el que se representan el funcionamiento de la solución.

2.1.1 Análisis y definición del problema

La problemática central reside en la ausencia de un sistema eficiente y accesible para la identificación precisa de las especies de aves en el Bosque Protector La Prosperina. Los métodos actuales de identificación dependen de la presencia física de un observador humano en el bosque, quien debe escuchar los cantos de las aves, lo cual limita la capacidad de monitoreo integral y periódico de las especies avícolas. La implementación de un sistema automatizado de reconocimiento de cantos de aves exige una extensa base de datos específica por sector o especie a clasificar, así como recursos informáticos, incluyendo estaciones de entrenamiento para el procesamiento de datos y la aplicación de algoritmos avanzados de procesamiento digital (DP). Además, se hace necesario establecer colaboraciones con expertos en ornitología para validar y mejorar de forma continua el sistema.

Figura 5

Diagrama de Actividades que representa el flujo del problema a solucionar.



El sistema de reconocimiento de audios de aves propuesto resolverá este problema al proporcionar una forma automática de identificar las especies de aves que habitan el bosque también deberá ser capaz de generar informes sobre la población de aves, lo que ayudará a los conservacionistas a monitorear la salud del bosque y a tomar medidas para proteger las especies en peligro de extinción.

2.1.2 Requerimientos

Estos requerimientos fueron levantados a través en reuniones con el cliente que tomaron un total de 4 horas, además de revisiones asíncronas de las historias de usuario en el que se contemplan lo siguiente:

Los requerimientos funcionales delinean las capacidades esenciales que el sistema debe ofrecer. En este contexto, se establecen requisitos clave, como la recopilación y almacenamiento de datos de audio de aves de diversas fuentes, el procesamiento de estos datos para la identificación de especies de aves, y la

generación de informes sobre la población de aves. Estos informes desempeñan un papel crucial en aplicaciones de seguimiento de la biodiversidad, investigación científica y monitoreo ambiental.

Por otro lado, los requerimientos no funcionales definen cómo debe operar el sistema en términos de disponibilidad, escalabilidad, seguridad, eficiencia en el uso de recursos y la capacidad de integrarse con otros sistemas de monitoreo ambiental. La alta disponibilidad del sistema, con una disponibilidad del 99,9%, asegura la continuidad en la recopilación de datos de audio. La escalabilidad es fundamental para gestionar el crecimiento constante de datos, mientras que la seguridad es esencial para proteger la privacidad de los datos de audio. Además, la eficiencia en el uso de recursos es un factor crucial para garantizar la sostenibilidad a largo plazo. Por último, la capacidad de integración con otros sistemas permite compartir datos valiosos con otras plataformas de monitoreo ambiental, fortaleciendo así su utilidad y alcance.

2.1.3 Usuarios y beneficiarios de la solución

Los usuarios y beneficiarios del sistema de reconocimiento de audios de aves incluyen; Científicos, Investigadores, Educadores y Público en general. Los científicos usarán el sistema para reconocer las especies de aves presentes en el bosque y seguir a su población. Los investigadores se beneficiarán de este sistema para analizar el comportamiento de las aves y su interacción con el entorno. Asimismo, los educadores utilizarán esta herramienta para instruir a los estudiantes acerca de las aves y su relevancia en el ecosistema. El público encontrará utilidad en este sistema para adquirir conocimientos sobre las aves residentes en el bosque y disfrutar de sus melodías.

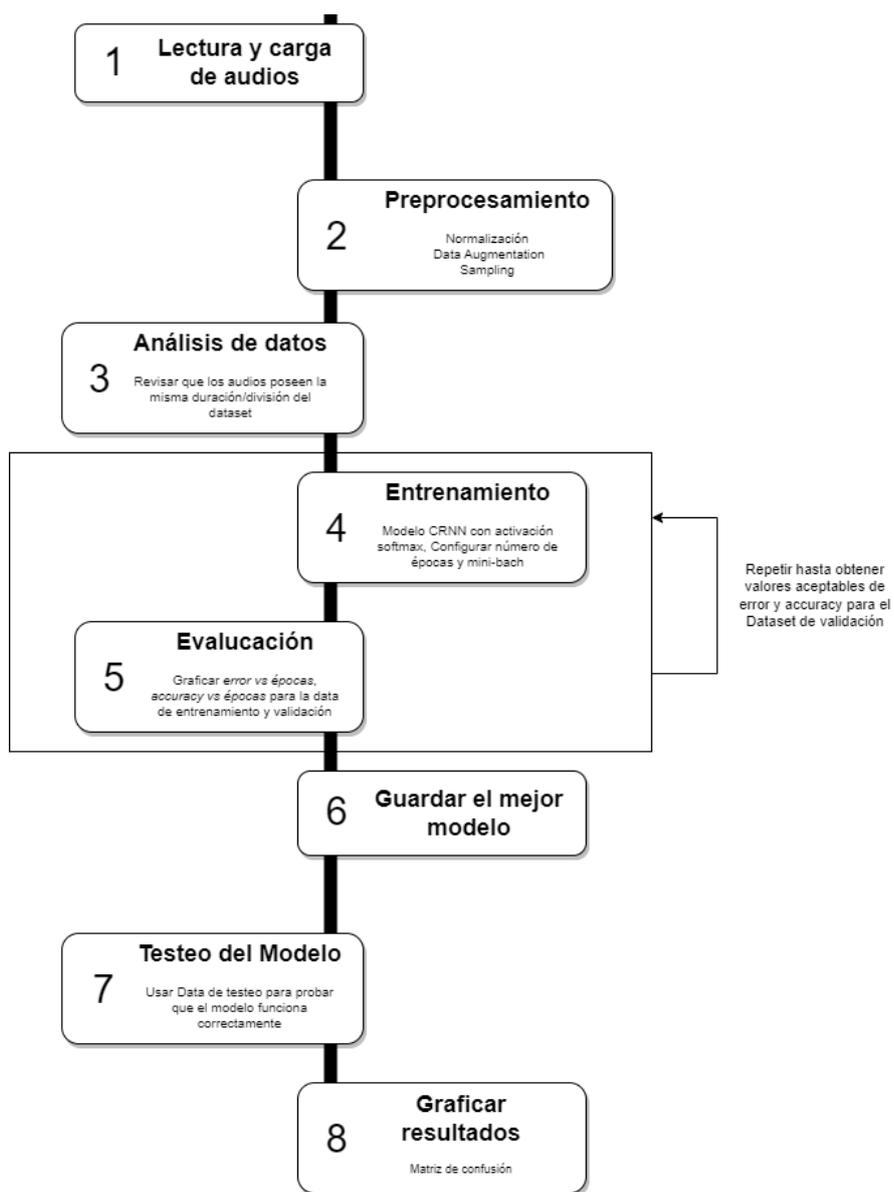
La implementación del sistema de reconocimiento de audios de aves tendrá un impacto positivo en la conservación de estas especies. Permitirá a científicos y conservacionistas identificar con precisión las diversas especies de aves presentes en el bosque, así como supervisar de cerca su población. Esta información será fundamental para que los conservacionistas tomen medidas para proteger a las especies en peligro de extinción y preservar su hábitat natural.

2.1.4 Diseño de la solución

Para diseñar la solución, se presentó el diagrama de bloques del proyecto, con componentes de lectura y carga de audios, preprocesamiento de la data, análisis, entrenamiento, evaluación y testeo del modelo para obtener los resultados gráficos.

Figura 6

Diagrama de bloques general del sistema de reconocimiento de aves.



2.1.5 Base de datos

La base de datos de audio de especies de aves en el Bosque de Protección "La Prosperina" utilizada para el entrenamiento del algoritmo fue una recopilación de grabaciones obtenidas de "xeno-canto.org", un sitio web gestionado por la Fundación Xeno-Canto que recopila grabaciones realizadas por grupos de investigación, individuos, fundaciones y otros, capturando especies de aves de diversas partes del mundo. Cada una de estas recopilaciones incluía el nombre científico del ave, la duración del audio, el autor de la grabación, la fecha, el país, la ubicación, la altitud y un identificador.

La mayoría de estas grabaciones consistían en ruido, sonidos de otras aves, lluvia, voces humanas, entre otros, lo que hacía más efectivo trabajar con clips de audio más cortos en lugar de grabaciones completas. Por lo tanto, las grabaciones de audio se cargaron en la plataforma de desarrollo de Edge Impulse para dividir manualmente los audios y obtener clips de audio de dos segundos con información importante, como el canto del ave.

2.1.6 Aumento de data

Dado que el número de archivos de sonido es relativamente pequeño, con un máximo de 70 archivos de sonido por clase en promedio, se necesitan métodos adicionales para evitar el sobreajuste. Además, el aumento de datos es uno de los elementos más cruciales para mejorar el rendimiento de generalización del sistema. Por lo tanto, se aplicaron dos métodos de aumento de datos.

2.1.6.1 Desplazamiento en el tiempo

Cada vez que se presente un ejemplo de entrenamiento a la red neuronal, se realiza una transformación de desplazamiento aleatorio en el tiempo. Si se trata de un espectrograma, lo divide en dos partes y se posiciona la segunda parte antes que la primera (desplazamiento de envolvente). De esta manera, se crea una esquina afilada donde el final de la segunda parte se encuentra con el comienzo de la primera, preservando toda la información. Esta técnica permite exponer a la red a variaciones irregulares en el

espectrograma y, lo más importante, enseñarle que los cantos y llamadas de aves pueden aparecer en cualquier momento en el audio, independientemente de la especie de ave.

2.1.6.2 Cambio de tono

La frecuencia del audio se modificó para alterar su tono. Para realizar esta transformación, se aplicaron técnicas de manipulación del espectro de frecuencia, como el desplazamiento de frecuencias hacia arriba o hacia abajo. Si se aumenta el tono, las frecuencias se desplazan hacia arriba, haciendo que el audio suene más agudo; mientras que, si se disminuye el tono, las frecuencias se desplazan hacia abajo, creando un sonido más profundo. Este proceso permite la generación de diferentes versiones del mismo audio, enriqueciendo el conjunto de datos de entrenamiento y ayudando al modelo a adaptarse a variaciones en el tono de las señales de entrada, mejorando así su capacidad de generalización en situaciones diversas.

2.1.7 Preprocesamiento

Generar características de entrada de alta calidad es crucial para el éxito de una red neuronal; por lo tanto, es esencial proporcionar una entrada limpia al modelo a través del preprocesamiento de datos de audio. Esto implica dos etapas principales. En primer lugar, se calcula el espectrograma de las componentes de señal y ruido. En segundo lugar, el espectrograma de cada componente se divide en segmentos de igual tamaño. En consecuencia, cada muestra del espectrograma de la señal es equivalente a un vector de características listo para ser utilizado como entrada de entrenamiento/prueba en el modelo de DL.

Para separar el archivo de sonido en señal y ruido, inicialmente se calculó el espectrograma de todo el archivo. La señal se somete a una Transformada de Fourier a Corto Plazo (STFT) utilizando la función 'reduce_noise' (manteniendo la señal de audio con un 75% de superposición). Sin embargo, la separación de señal-ruido se desvía de este procedimiento estándar; por lo tanto, cada elemento se divide por el valor máximo, asegurando que todos los valores estén dentro del rango [0, 1]. Una vez obtenido el espectrograma, se identificaron los intervalos de señal-ruido.

2.1.8 Arquitectura

La estructura de nuestra red se centró en resolver el desafío de reconocer especies de aves cuando no disponíamos de etiquetas claras. La red tomó grabaciones de sonidos de aves, sin información de qué especie pertenecía a cada grabación. Esto reflejaba la realidad de trabajar con datos no etiquetados, donde la tarea era identificar automáticamente las especies sin guía previa.

En la fase inicial, se utilizó un autoencoder variacional para extraer características fundamentales de los cantos de aves. El codificador mapeó los sonidos a un espacio latente donde se introdujeron variaciones estocásticas, proporcionando adaptabilidad. El decodificador reconstruyó los sonidos a partir de estas representaciones latentes.

Después de la fase VAE, se incorporó un módulo de clustering que asignó diferentes grabaciones a grupos o "clusters". Esto se hizo sin saber de antemano qué especie representaba cada grabación. La red aprendió patrones similares y agrupó naturalmente grabaciones que compartían características acústicas comunes.

La salida final reveló a qué grupo pertenecía cada grabación de ave, sin la necesidad de etiquetas previas. Este enfoque fue valioso para identificar patrones y agrupar cantos de aves de manera significativa, incluso cuando no teníamos información detallada sobre las especies en las grabaciones.

Al unir la capacidad de aprendizaje del VAE con la capacidad de agrupación del módulo de clustering, nuestra red pudo descubrir automáticamente patrones y similitudes entre los cantos de aves sin depender de etiquetas específicas de especies. Esto proporcionó una herramienta poderosa para la identificación y clasificación de especies de aves en situaciones en las que la información de referencia era limitada o inexistente.

2.1.9 Detección de anomalías

Se utilizó el modelo de k-vecinos más cercanos, técnica de detección de anomalías que se fundamenta en la identificación de las muestras que exhiben una mayor distancia con respecto al conjunto de datos predominante. En el contexto de la clasificación de vocalizaciones de aves, esto implica identificar aquellas que difieren significativa o distantes con respecto a las de aves conocidas.

Para implementar este enfoque, se requiere, primero, contar con datos que incluya las vocalizaciones de varias especies de aves (correspondiente a la categoría normal) y otras formas de sonidos o grabaciones de audio que no corresponden a vocalizaciones de aves (pertenecientes a la categoría de anomalías). Después, el modelo de k-vecinos más cercanos se entrena usando las representaciones de las muestras de audio en un espacio de características apropiado.

Durante el proceso de detección, cuando se presenta una nueva vocalización de ave al modelo, este procede a calcular la distancia o similitud entre la vocalización en cuestión y sus vecinos más cercanos dentro del espacio de características. Si la vocalización se encuentra significativamente distante o presenta una menor similitud con respecto a las muestras de vocalizaciones de aves conocidas, puede que se clasifique como una anomalía o como un sonido sin la categoría normal de vocalizaciones de aves.

2.1.10 Librerías

Para el desarrollo del algoritmo se utilizó el lenguaje de programación de Python en su versión 3.9, ya que es un lenguaje que proporciona diferentes herramientas y librerías para la programación de modelos de inteligencia artificial.

A continuación, se especifican las librerías que se manejaron:

Numpy. Es una librería de Python para el procesamiento de array. Contiene una gran colección de funciones que permiten realizar cálculos matemáticos complejos sobre arrays multidimensionales. La

funcionalidad que implementa permite realizar operaciones lógicas, redimensiones, búsquedas y aplicar estadísticas entre otras muchas.

Matplotlib. Librería de Python que permite la visualización de datos de cualquier índole e incluso visualizaciones en 3D.

Keras. La biblioteca Keras ofrece una interfaz de alto nivel que permite acceder fácilmente a las funciones de TensorFlow, una herramienta utilizada para desarrollar modelos de deep learning a un nivel más bajo. Al combinar ambas bibliotecas, es posible crear redes neuronales complejas de manera intuitiva, definiendo cada elemento de la red, incluyendo su arquitectura en capas, las funciones de activación y los criterios de evaluación, entre otras características.

Librosa Python. Librería de código abierto que permite a los usuarios trabajar con archivos de audios que parten desde la carga y visualización hasta extraer funciones del audio para recolectar características como la tasa de cruce por cero, coeficientes cepstrales de frecuencia Mel (MFCC), entre otros.

2.1.11 Interfaz web

Para facilitar a los conservacionistas la revisión de las predicciones, se desarrolló una interfaz funcional de página web. Esta interfaz permite a los usuarios cargar grabaciones de audio en formato .wav para su predicción. Cuando el modelo prevé, se muestran detalles esenciales del audio, como la duración, el nombre, la vista previa del audio y la lista de aves predichas junto con sus respectivas imágenes.

Además, se presenta una serie de tiempo, segmentada en intervalos de predicción de 8 segundos, donde los picos representan intervalos de tiempo detectados. Las detecciones ocurren en intervalos de 2 segundos, lo que significa que cada 8 segundos se aplica un algoritmo de voto mayoritario de Boyer-Moore [29]. En este algoritmo, el valor que ocurre con más frecuencia se convierte en la predicción para ese segmento de 8 segundos, ya sea una detección de ave o no. Por ejemplo, si se registran cuatro predicciones de aves

diferentes dentro de un intervalo de 8 segundos, podría indicar un ave no entrenada, y se marca como una anomalía.

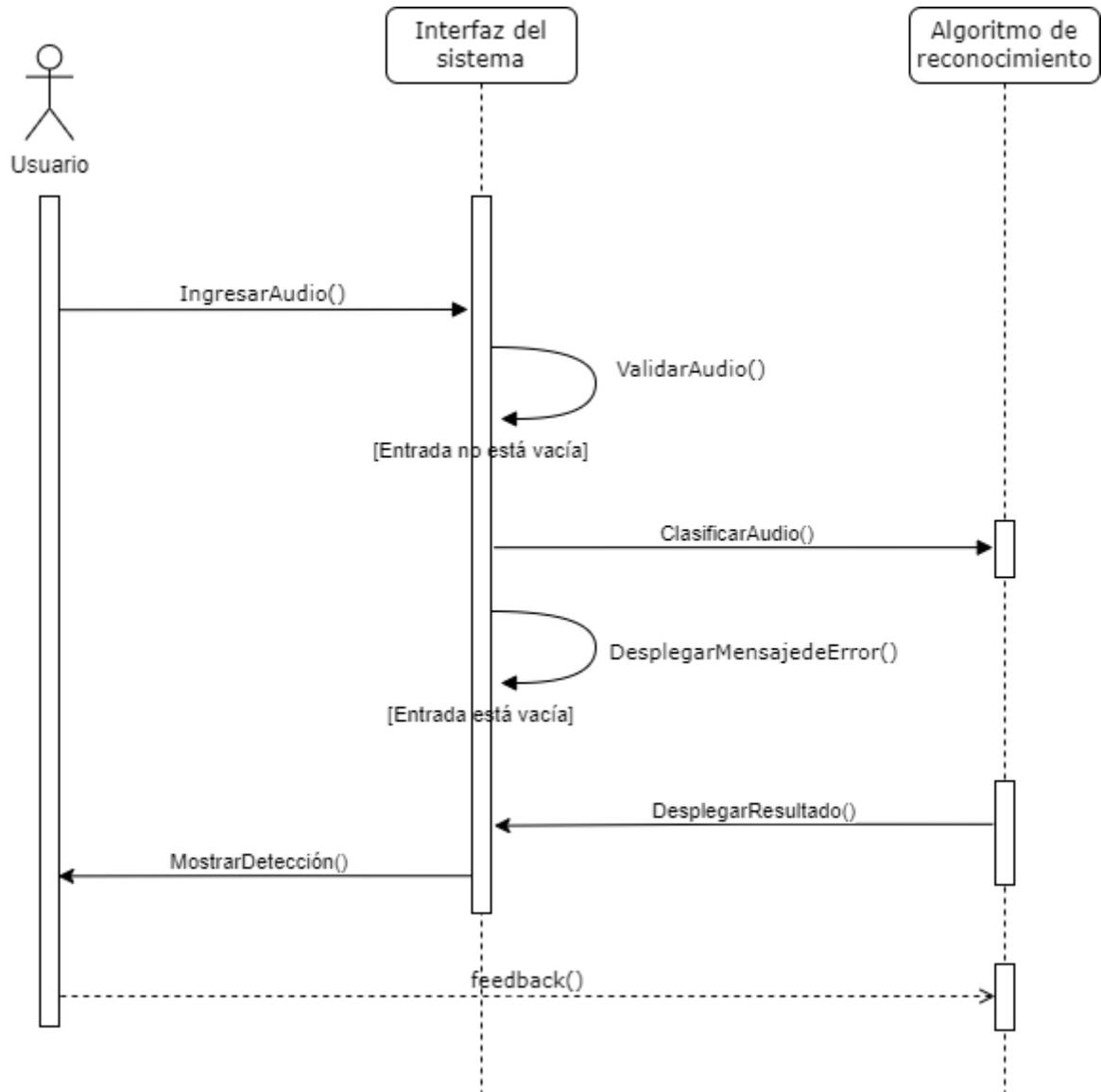
Por último, la interfaz no solo facilita la identificación de especies avícolas, sino que también brinda la conveniente función de descargar informes detallados. Estos informes, disponibles en formato .csv, proporcionan datos precisos sobre las especies detectadas en el audio.

2.1.12 Diagramas

Se agregó un diagrama de secuencia en el que se expresan las interacciones entre el usuario y el sistema que se manejaron en la solución completa con la interfaz de usuario en conjunto con el algoritmo de Deep Learning, desde la entrada de audio hasta el despliegue del resultado y retroalimentación de este.

Figura 7

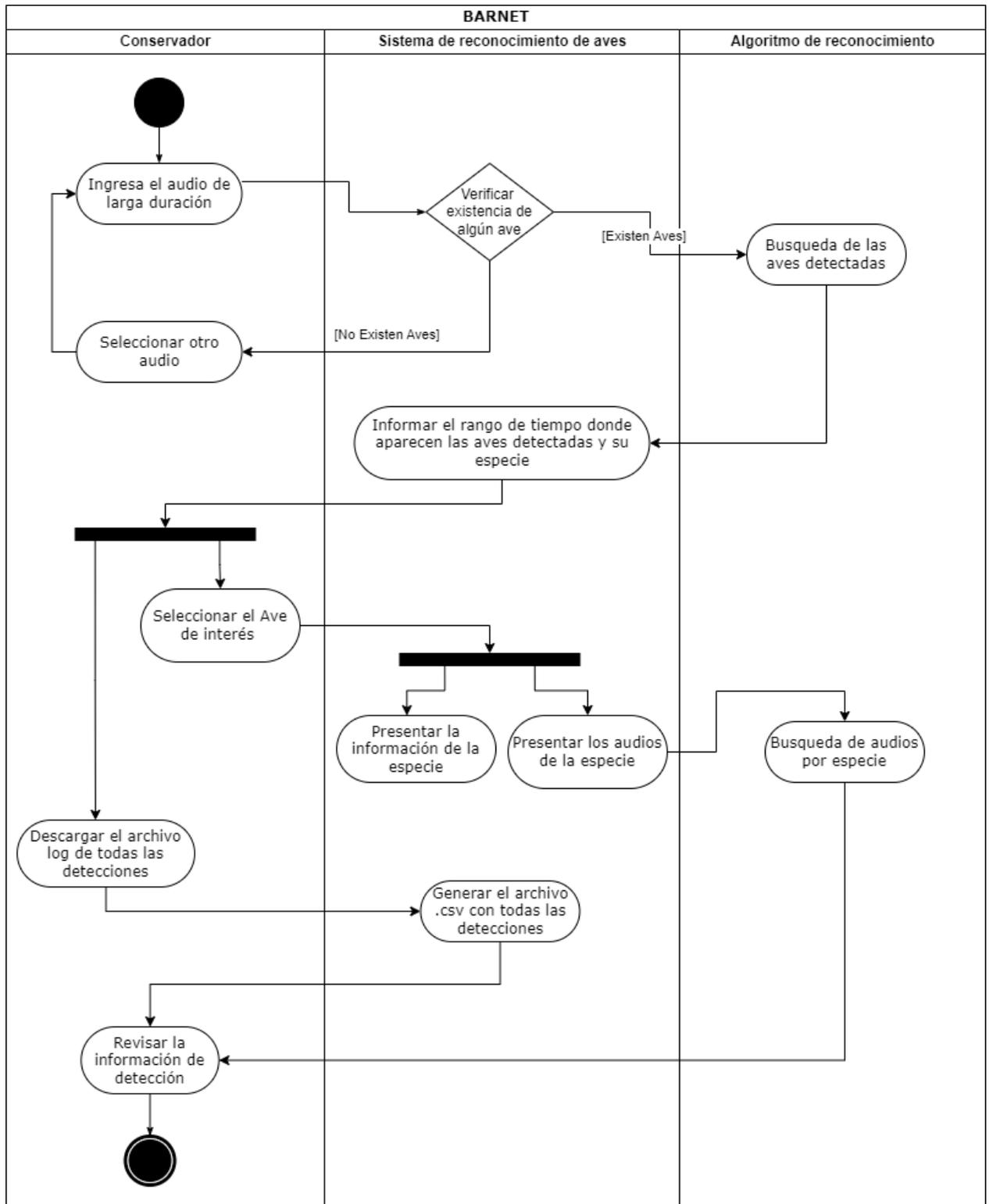
Diagrama de secuencia.



Además, se graficó un diagrama de actividades que refleja las acciones de un usuario específico, un conservacionista enfocado a las aves, y del sistema de reconocimiento de audios de aves realizadas.

Figura 8

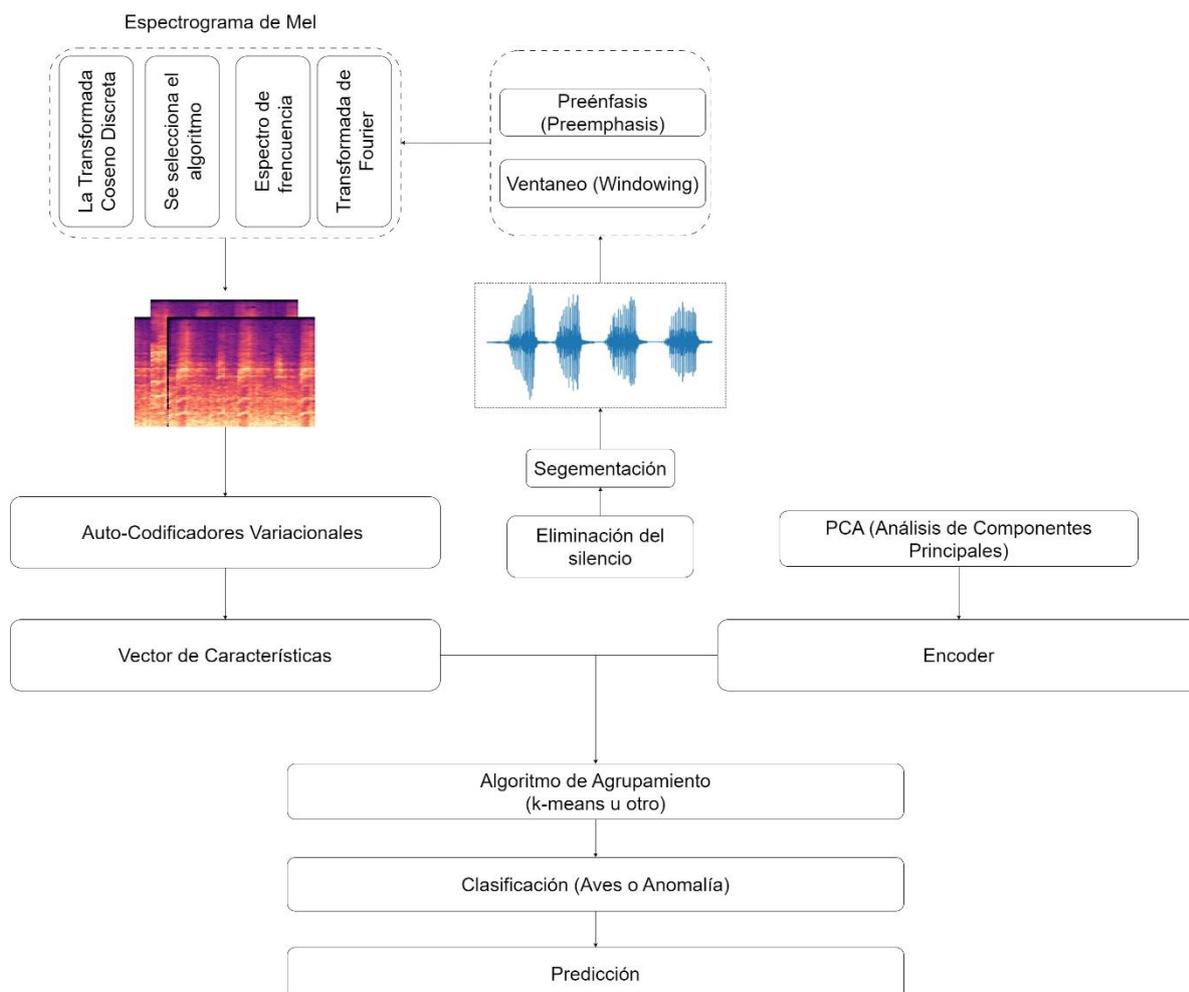
Diagrama de Actividades que representa el flujo de la solución.



Por último, se creó un diagrama que representa el flujo de procesamiento para el reconocimiento de aves mediante Deep learning. Este diagrama detalla la secuencia de pasos, desde la entrada de la grabación de audio hasta la clasificación final, incluyendo el preprocesamiento de audio, la extracción de características utilizando el modelo pre-entrenado de IA, y el uso de un algoritmo de agrupamiento para identificar la especie de ave o detectar anomalías.

Figura 9

Diagrama que representa el flujo de la extracción de características hasta la predicción de la Especie a Clasificar.



CAPÍTULO 3

3.1 Plan de evaluación

En la (Figura 10) se muestra a una visión general del plan de evaluación de la aplicación web que alberga el modelo entrenado para la identificación de especies de aves en peligro de extinción en los bosques secos. Este análisis se centró en tres metas principales del producto. La primera de ellas se enfocó en el porcentaje de precisión del modelo en la clasificación de especies. La segunda meta abordó el porcentaje de reducción de la carga laboral asociada a las tareas de clasificación de especies y cálculos poblacionales. Finalmente, la tercera meta evaluó el porcentaje de error en las predicciones del modelo.

El estudio contó con la participación de aproximadamente 30 usuarios, caracterizados por dos perfiles distintos. En primer lugar, se involucraron visitantes del Bosque Protector La Prosperina, que representan el público general y, por ende, usuarios potenciales de la aplicación web. En segundo lugar, se incorporaron 25 conservacionistas y profesionales con carreras afines, quienes poseen experiencia y conocimientos especializados en la clasificación de aves.

Figura 10

Plan de evaluación de la aplicación.

Nº	Meta del Producto	Ideas de Evaluación	Actividades de Evaluación	Perfil del Usuario	Cantidad de Usuarios	Duración de la Actividad	Fecha
1	Identificar especies de aves en peligro de extinción que viven en los bosques secos con una precisión del 95%	Matriz de confusión clase de ave: porcentaje de aciertos Feedback por parte de los expertos sobre la predicción presentada: comentarios positivos sobre la predicción. Porcentaje de aciertos entre el conjunto de entrenamiento y el de testeo	Se crea una matriz de confusión en el que se pueda apreciar el porcentaje de acierto por cada clase para evaluar el modelo, en la aplicación el usuario revisa la información de la predicción.	Visitante del bosque protector la prosperina	30	15 días	3 de diciembre
2	La aplicación reduce la carga laboral en la clasificación de especies en audios de gran extensión en un 70%	Comparación entre las jornadas laborales a la hora de hacer cálculos antes y después del uso de la app: porcentaje de reducción de jornada Se contabilizaría el tiempo que le toma al conservacionista Tiempo dedicado en encontrar especies en audios de gran extensión antes y después del uso de la app: porcentaje de tiempo dedicado	Los conservacionistas suben los audios a la plataforma y revisan las especies que fueron encontradas y mediante el gráfico de especies encontradas o con el reporte descargable por periodo de tiempo van haciendo cálculos.	Conservacionista	10	7 días	18 de diciembre
3	Porcentaje de error en el reconocimiento de especies de aves menor al 3%	Gráfica de porcentaje de error: porcentaje de error general de las predicciones Feedback por parte de los expertos sobre la predicción presentada: comentarios negativos sobre la predicción. Porcentaje de error entre el conjunto de entrenamiento y el de testeo	Los conservacionistas suben los audios a la plataforma y revisan las especies que fueron encontradas para que corroboren si hay errores en la predicción como especies mal etiquetadas (ej: canto de un carpintero de guayaquil etiquetado como Amazona Frentirroja)	Conservacionista	15	9 días	4 de enero de 2024

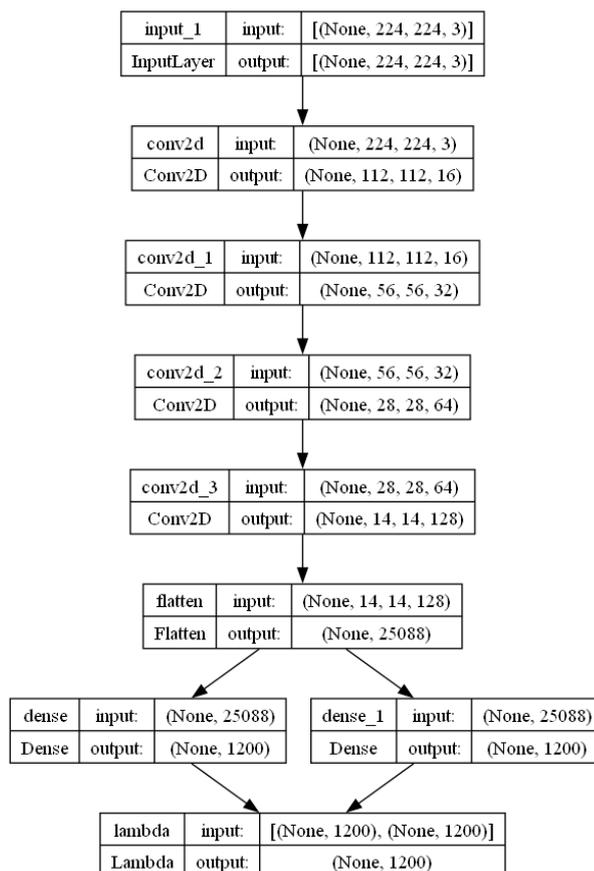
3.2 Resultados y análisis

En la estructura integral de la VAE desarrollado para la detección de cantos de aves, se implementó un codificador (encoder) y un decodificador (decoder).

El codificador se diseñó para capturar las características fundamentales usando capas convolucionales bidimensionales (Conv2D), con tres capas sucesivas con 16, 32, 64 y 128 filtros, respectivamente. Estas capas, con activación ReLU y operando en un contexto espacial definido por filtros de tamaño (3, 3), facilitaron la extracción de información significativa. Posteriormente, la salida de las capas convolucionales fue aplanada mediante una capa Flatten, y dos capas densas fueron empleadas para obtener tanto la media (z_mean) como el logaritmo de la varianza (z_log_var) de la distribución latente. La forma resultante antes de aplanar se registró para su uso en las etapas posteriores del modelo.

Figura 11

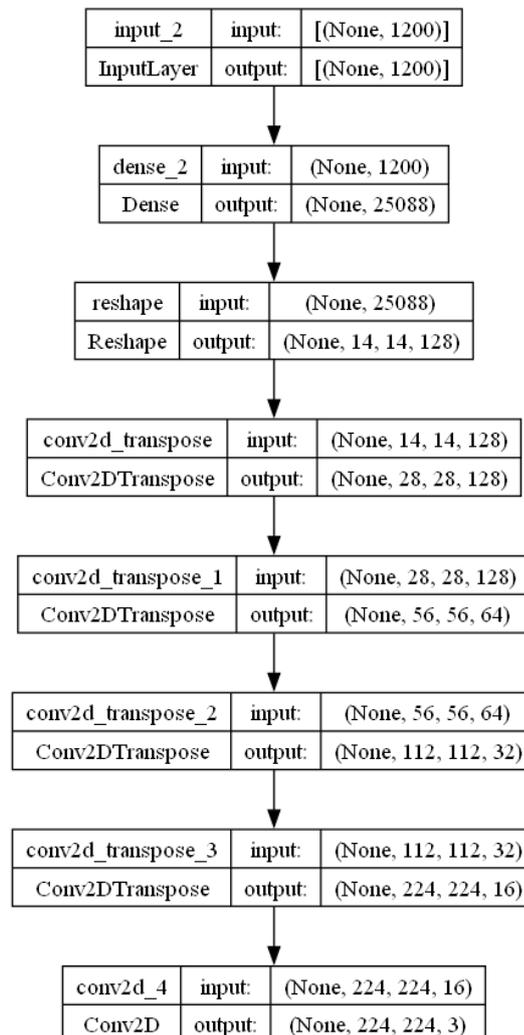
Encoder del VAE



En cuanto al decodificador, se diseñó para reconstruir la representación latente (z) de manera precisa. Inicialmente, una capa densa con activación ReLU transformó la entrada proveniente del espacio latente a un formato compatible con la forma antes del aplanamiento, que fue recuperada durante la codificación. A continuación, una operación de remodelado (Reshape) ajustó las dimensiones de la salida para adaptarse a las características originales antes del aplanamiento. El proceso de reconstrucción continuó con el uso de capas de convolución transpuesta (Conv2DTranspose) con activación ReLU y relleno 'same', implementadas en cuatro capas sucesivas con 128, 64, 32 y 16 filtros respectivamente. Finalmente, una última capa de convolución con activación sigmoide y relleno 'same' generó la salida final del modelo, compuesta por tres canales, representando así la reconstrucción completa del espacio de entrada original.

Figura 12

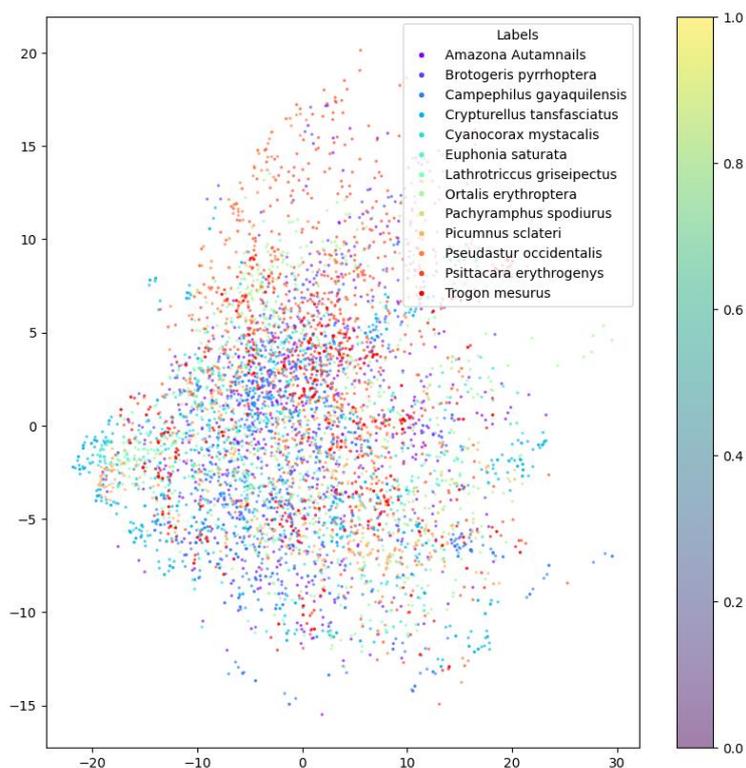
Decoder del VAE



Tras completar el entrenamiento de la VAE con los espectrogramas de los audios de las aves, se procedió a analizar el espacio latente resultante. Este análisis se materializa en el gráfico de dispersión del espacio latente, que representa visualmente las relaciones entre las características latentes aprendidas durante la codificación. En este gráfico, cada punto corresponde a una representación única en el espacio latente para un espectrograma de audio específico. La dispersión y la distribución de estos puntos revelan patrones y agrupamientos que proporcionan información valiosa sobre la estructura subyacente de los datos. Este análisis del espacio latente a través del gráfico de dispersión no solo valida la capacidad del modelo para aprender representaciones significativas, sino que también ofrece percepciones clave sobre la variabilidad y la correlación entre las características latentes extraídas de los espectrogramas de las vocalizaciones de las aves.

Figura 13

Gráfico de dispersión / Espacio latente



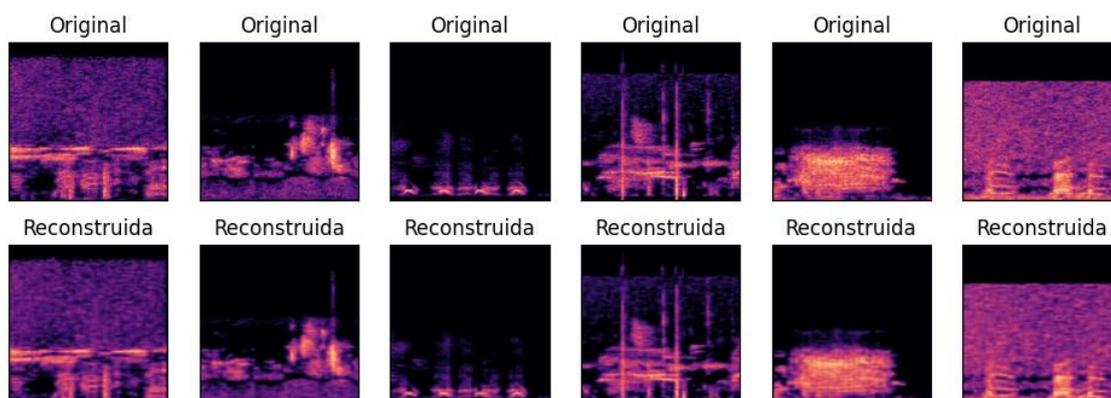
Con relación a la reconstrucción del audio, se puede apreciar en la (Figura 13) como se logró una reconstrucción prácticamente perfecta del espectrograma. Este resultado sugiere que el vector en el espacio latente sirve como una representación eficaz de las características fundamentales de un espectrograma. En

otras palabras, este vector captura de manera significativa las propiedades esenciales que definen la estructura y contenido del espectrograma original.

Estos resultados se respaldan cuantitativamente mediante los valores obtenidos durante el entrenamiento del VAE, donde se registró un `val_loss` de (290664.21), desglosado en `val_kl_loss` (4936.73) y `val_reconstruction_loss` (285.72). La pérdida de divergencia de Kullback-Leibler refleja la eficacia del modelo en aprender una representación comprimida en el espacio latente, mientras que la pérdida de reconstrucción cuantifica la precisión de la recuperación de datos originales. Estos indicadores subrayan la robustez del VAE en la tarea de reconstrucción, respaldando la conclusión de que la representación latente facilita una reconstrucción precisa del espectrograma.

Figura 14

Reconstrucción de la imagen del Espectrograma.

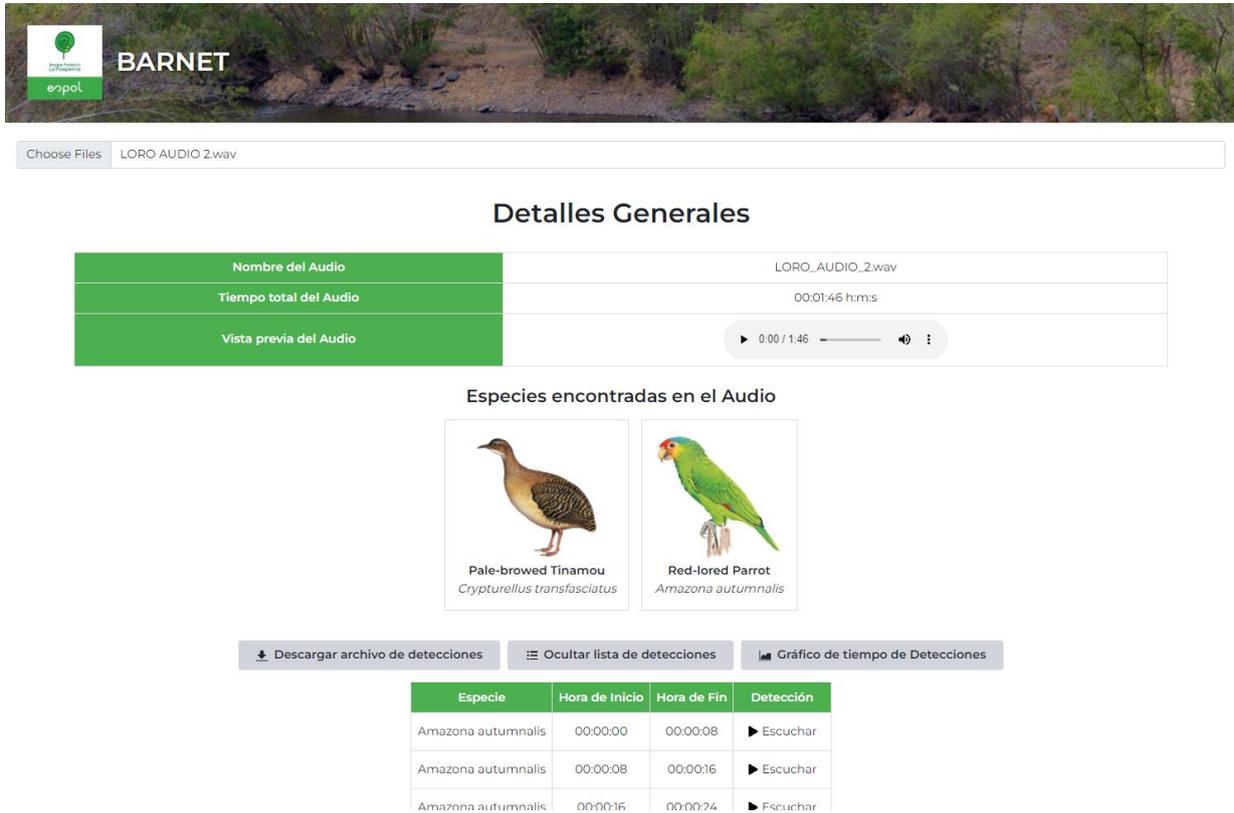


3.3 Interfaz Web

Con la implementación local de la interfaz de reconocimiento de aves (Fig. 15), los conservacionistas e investigadores pueden identificar eficazmente las aves.

Figura 15

Implementación local de la interfaz.



El gráfico de series temporales (Fig. 16) muestra las detecciones de aves, marcando las instancias visibles donde se ha detectado la presencia de aves y aquellas donde no ha sido detectada. Esta representación visual permite observar la evolución de las detecciones a lo largo del tiempo. Además, proporciona una perspectiva valiosa para evaluar la precisión de las detecciones en relación con la verificación de audio.

Figura 16

Serie de tiempo de predicciones.



Por último, la opción para descargar los informes sobre las predicciones de las grabaciones (Tabla. 2), son clave para futuros estudios que podamos realizar. Estos informes no solo nos ayudan a documentar de manera sistemática los resultados, sino que también son esenciales para hacer análisis detallados y evaluar críticamente la información en investigaciones posteriores.

Tabla 2

Ejemplo de los campos que posee el informe de predicciones .CSV

Nombre_Archivo	Nombre_Común	Nombre_Científico	Tiempo_Inicio	Tiempo_Fin
Audio_de_la_Plana_chac halaca_.HD.wav	Not Detected	Not Detected	0:00:00	0:00:08
Audio_de_la_Plana_chac halaca_.HD.wav	Chachalaca Cabecirrufa	Ortalis erythroptera	0:00:08	0:00:16
Audio_de_la_Plana_chac halaca_.HD.wav	Chachalaca Cabecirrufa	Ortalis erythroptera	0:00:16	0:00:24

3.4 Análisis de Costos y viabilidad

Los costos asociados con el desarrollo del proyecto se mantuvieron en niveles bajos, en gran parte gracias al uso de herramientas y bibliotecas de código abierto provenientes del entorno de programación Python. En particular, se optó por usar versiones gratuitas de herramientas como Edge Impulse, pese a tener opciones pagas, para crear el conjunto de datos etiquetados de aves. Además, los archivos de audio empleados en la investigación se obtuvieron de manera pública y sin costo alguno a través del sitio web de la fundación Xeno-Canto y la biblioteca Macaulay, posibilitando su descarga y utilización sin generar gastos adicionales.

Es imperativo resaltar que, a pesar de que la mayoría de las herramientas y recursos utilizados no conllevaron costos directos, el proceso de entrenamiento efectivo del modelo de reconocimiento de audio de aves puede requerir el uso de una Unidad de Procesamiento Gráfico (GPU).

Desde una perspectiva de viabilidad, el proyecto se considera económicamente sostenible y tecnológicamente factible, dado que no fue necesario realizar inversiones monetarias significativas. No

obstante, se debe tener en cuenta que, para un entrenamiento eficaz del modelo, es esencial evaluar la disponibilidad de recursos GPU.

3.5 Aporte Académico-Científico

Este proyecto contribuye mucho al ámbito académico-científico, en especial en el estudio de la detección de cantos de aves. La implementación de un modelo de deep learning diseñado específicamente para la identificación de aves a partir de grabaciones de audios extensos introduce nuevos enfoques y metodologías innovadoras en el campo de la ornitología, que se ocupa del estudio de las aves.

El Bosque Protector La Prosperina se beneficiará con la incorporación de esta herramienta, que aporta a la monitorización de especies y al estudio de la biodiversidad identificando automatizada de cantos de aves. Este aporte tecnológico promete mejorar la eficiencia en la recopilación de datos en el terreno y también aumentar la precisión de la misma, promoviendo un enfoque más sostenible para conservar la fauna aviar en la mencionada área.

En resumen, la implementación de este modelo de deep learning no solo enriquece el conocimiento científico en el ámbito de la ornitología, sino que también ofrece una herramienta valiosa para la gestión y preservación de la biodiversidad en el Bosque Protector La Prosperina.

3.6 Publicación de artículo científico

Este trabajo es una continuación de un artículo científico publicado llamado “Towards Birds Conservation in Dry Forest Ecosystems through Audio Recognition via Deep Learning” para la Internacional Conference on Information and Computer Technologies (ICICT) en su edición 2024. El artículo se presentó en línea para el 21 de febrero 2024. [31]

CAPÍTULO 4

4.1 Conclusiones y recomendaciones

4.1.1 Conclusiones

- La aplicación del algoritmo de reconocimiento de cantos permitió identificar con precisión especies de aves presentes en el Bosque Seco, priorizando aquellas en algún grado de amenaza. Este logro respalda el objetivo de conservación al proporcionar información detallada sobre poblaciones críticas.
- El algoritmo facilitó la recopilación eficiente de registros de aves basados en sus cantos. Esto no solo simplificó el proceso de seguimiento, sino que también mejoró la calidad y cantidad de datos disponibles para futuras investigaciones y planes de conservación.
- El diseño y desarrollo del algoritmo demostraron ser robustos para el reconocimiento de cantos en audios extensos. La estructura bien definida permitió una adaptabilidad efectiva a diferentes duraciones y contextos, cumpliendo con el objetivo de diseñar un algoritmo eficiente.
- La implementación de técnicas de aprendizaje profundo, con el uso de auto decodificadores variables, contribuyó al desarrollo del algoritmo de reconocimiento de cantos. Esta aproximación avanzada garantizó la precisión y generalización del modelo, cumpliendo así con el objetivo de aplicar técnicas de vanguardia.
- La evaluación del desempeño del algoritmo proporcionó resultados prometedores, respaldando su utilidad en el apoyo a planes de conservación. La precisión y confiabilidad del reconocimiento de especies refuerzan la viabilidad de utilizar esta tecnología en iniciativas de conservación aviar.

4.1.2 Recomendaciones

- Se recomienda continuar la expansión del conjunto de datos con grabaciones de una variedad más amplia de especies y contextos acústicos. Esto fortalecerá la capacidad del modelo para generalizar y adaptarse a nuevas situaciones.
- La interfaz web podría beneficiarse de mejoras para ofrecer una experiencia más intuitiva y amigable. La inclusión de funciones adicionales, como la carga y etiquetado de nuevas grabaciones, podría hacerla más versátil y completa.

- Dada la dinámica de cambio en los entornos naturales, se sugiere una evaluación continua y actualización del modelo. Incorporar datos recientes garantizará que el modelo siga siendo relevante y preciso a medida que las poblaciones de aves evolucionan.
- Fomentar la colaboración con la comunidad científica y ornitológica permitirá una retroalimentación valiosa. Compartir resultados y metodologías contribuirá al avance general en la identificación y conservación de ave

Referencias

- [1] M. A. Tabur and Y. Ayvaz, "Ecological Importance of Birds," Suleyman Demirel University, Science and Art Faculty, Biology Department, 32260 Isparta, Turkey.
- [2] BirdLife International, "IUCN Red List for birds," 2019. [Online]. datazone.birdlife.org/species/search.
- [3] Unión Internacional para la Conservación de la Naturaleza, International Union for Conservation of Nature, 2023. [Online]. www.iucnredlist.org/es.
- [4] J. Freile, T. Santander, G. Jiménez-Uzcátegui, L. Carrasco, D. Cisneros-Heredia, E. Guevara, M. Sánchez-Nivicela y B. Tinoco, "Red List of the Birds of Ecuador," 2019.
- [5] BirdLife International, "Country profile: Ecuador," 2023, [Online]. datazone.birdlife.org/country/ecuador.
- [6] ESPOL, Update Of The Management Plan For La Prosperina Protected Forest. 2020.
- [7] S. Fagerlund, "Bird Species Recognition Using Support Vector Machines", EURASIP J. Adv, Signal Process, 038637, 2007. doi.org/10.1155/2007/38637
- [8] T. Scott Brandes, "Automated sound recording and analysis techniques for bird surveys and conservation," Bird Conservation International, vol. 18, no. S1, pp. S163-S173, 2008, doi: 10.1017/S0959270908000415.
- [9] P. Amlathe, "Standard Machine Learning Techniques in Audio Beehive Monitoring: Classification of Audio Samples with Logistic Regression, K-Nearest Neighbor, Random Forest and Support Vector Machine", Master's Thesis, Utah State University, May 2018. doi.org/10.26076/7e6c-25b1
- [10] Xeno-canto Foundation, "xeno-canto Sharing wildlife sounds from around the world," 2023. [Online]. xeno-canto.org.
- [11] American Bird Conservancy(2020). Importancia de las Aves [Online]. Disponible en: <https://storymaps.arcgis.com/stories/576df9f288bb40da84584b1dd8e7802b>
- [12] Santiago, G. V. (2017) La Ciencia Acústica en los Ambientes Inteligentes, Latin American Journal of Computing [en prensa].

- [13] G. Gallo, "La bioacústica y la ecoacústica al servicio de la valoración de los ecosistemas", Inecol.mx. [En línea]. Disponible en: <https://www.inecol.mx/inecol/index.php/es/ct-menu-item-25/ct-menu-item-27/17-ciencia-hoy/1891-la-bioacustica-y-la-ecoacustica-al-servicio-de-la-valoracion-de-los-ecosistemas>.
- [14] "Rainforest connection - stopping illegal logging & protecting wildlife", Rfcx.org. [En línea]. Disponible en: <https://rfcx.org/impact>.
- [15] El universo, "Un laboratorio natural en Bosque Prosperina", ESPOL, 22-jun-2018. [En línea]. Disponible en: https://www.espol.edu.ec/sites/default/files/docs_escrabe/Un%20laboratorio%20natural%20en%20Bosque%20Prosperina.pdf.
- [16] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. In Proceedings of the 2018 World Wide Web Conference (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 689–698. <https://doi.org/10.1145/3178876.3186150>
- [17] Gatopoulos I, Tomczak JM. Self-Supervised Variational Auto-Encoders. Entropy. 2021; 23(6):747. <https://doi.org/10.3390/e23060747>
- [18] J. de la Torre, "AUTOCODIFICADORES VARIACIONALES (VAE) FUNDAMENTOS TEÓRICOS Y APLICACIONES", en Arxiv.org, 2023.
- [19] Qingzhong Liu. (2017). A Comparison Study to Identify Birds Species Based on Bird Song Signals [Online]. Disponible en: https://www.researchgate.net/publication/319485395_A_Comparison_Study_to_Identify_Birds_Species_Based_on_Bird_Song_Signals
- [20] L. Wyse, "Audio spectrogram representations for processing with convolutional neural networks", arXiv [cs.SD], 2017.
- [21] B. Zhang, J. Leitner, and S. Thornton, "Audio Recognition using Mel Spectrograms and Convolution Neural Networks," Dept. of Electrical and Computer Engineering, University of California, San Diego.

- [22] T. Kaneko, K. Tanaka, H. Kameoka, and S. Seki, "iSTFTNet: Fast and lightweight mel-spectrogram vocoder incorporating inverse short-time Fourier transform," in ICASSP (International Conference on Acoustics, Speech and Signal Processing) 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2022, pp. 6207-6211
- [23] A. Schrage, "Seeing Song: An Interview with a Partially Deaf Ornithologist Who Studies Bird Song," in *Tough Little Birds*, January 26, 2015.
- [24] Lei Zhang, Haifeng Wang, and Yu-Gang Jiang (2018). A Comparative Study of Noise Reduction Techniques for Bird Song Recognition [Online]. Disponible en: https://www.carolinabrum.com/publications/cbrum_birds.pdf
- [25] Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann. "Audio Based Bird Species Identification using Deep Learning Techniques". [Online]. Disponible en: https://www.researchgate.net/publication/338422871_Audio-based_Bird_Species_Identification_with_Deep_Convolutional_Neural_Networks
- [26] Devopedia. 2021. "Audio Feature Extraction." Version 8, May 23. [Online]. Disponible en: <https://devopedia.org/audio-feature-extraction>
- [27] K. Darras, P. Batáry, B. J. Furnas, I. Grass, Y. A. Mulyani, y T. Tschardtke, "Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide", *Ecol. Appl.*, vol. 29, núm. 6, 2019.
- [28] S. Zhang, Y. Gao, J. Cai, H. Yang, Q. Zhao, and F. Pan, "A Novel Bird Sound Recognition Method Based on Multifeature Fusion and a Transformer Encoder," *Sensors*, vol. 23, no. 19, 2023, p. 8099. <https://doi.org/10.3390/s23198099>.
- [29] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck, "BirdNET: A deep learning solution for avian diversity monitoring," *Ecological Informatics*, vol. 61, 2021, p. 101236. ISSN 1574-9541. <https://doi.org/10.1016/j.ecoinf.2021.101236>.
- [30] R. S. Boyer and J. S. Moore, "MJRTY—a fast majority vote algorithm," in *Automated reasoning: essays in honor of Woody Bledsoe*, Springer Netherlands, 1991, pp. 105-117.

- [31]T. E. Rodríguez,A. B. Guilindro, P. M. Piedrahita, M. A. Realpe, “Towards Birds Conservation in Dry Forest Ecosystems through Audio Recognition via Deep Learning”, en Internacional Conference on Information and Computer Technologies, Febrero 2024