

Sistema de procesamiento masivo de datos bioacústicos para el monitoreo automático de aves mediante Deep Learning

PROBLEMA

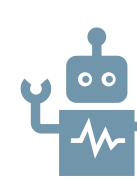
El uso de grabadoras automáticas en la naturaleza genera miles de horas de audio que son difíciles y costosas de analizar manualmente. La identificación experta de cantos requiere demasiado tiempo y retrasa el monitoreo continuo de la biodiversidad. Además, el ruido ambiental causado por viento, lluvia e insectos dificulta reconocer con precisión las vocalizaciones. Herramientas globales como BirdNET pueden perder precisión en ecosistemas locales como el Bosque Protector Chongón-Colonche, generando falsos positivos.

OBJETIVO GENERAL

Desarrollar e implementar un sistema automático basado en Inteligencia Artificial (Deep Learning) para procesar grandes volúmenes de grabaciones de audio e identificar especies de aves en entornos naturales de forma rápida, precisa y no invasiva.

PROPUESTA

Una herramienta que permite procesar masivamente grabaciones bioacústicas para detectar de forma automática cantos y llamadas de aves mediante técnicas de Inteligencia Artificial.

- 
 - El módulo de preprocesamiento filtra el ruido ambiental y transforma las grabaciones en espectrogramas Mel para la detección
- 
 - El módulo de detección emplea YOLO26 para localizar automáticamente las vocalizaciones presentes en los espectrogramas.
- 
 - La solución incorpora procesamiento paralelo, mediante múltiples trabajadores (workers) que procesan simultáneamente diferentes grabaciones, además de una interfaz web en Streamlit para gestionar y visualizar resultados.
- 
 - El sistema genera detecciones, recortes de las vocalizaciones identificadas y sus métricas asociadas, facilitando la revisión posterior de las grabaciones.

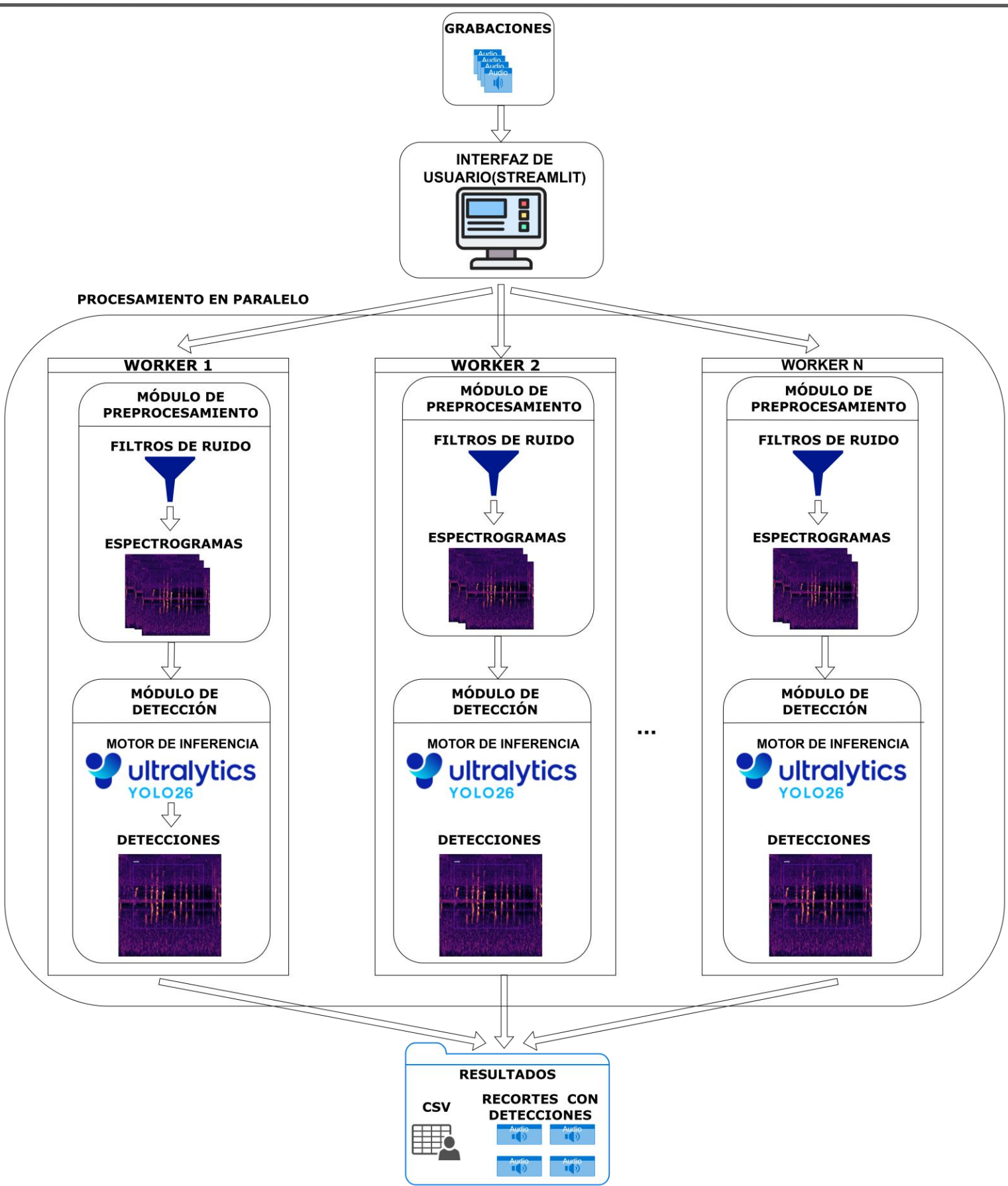
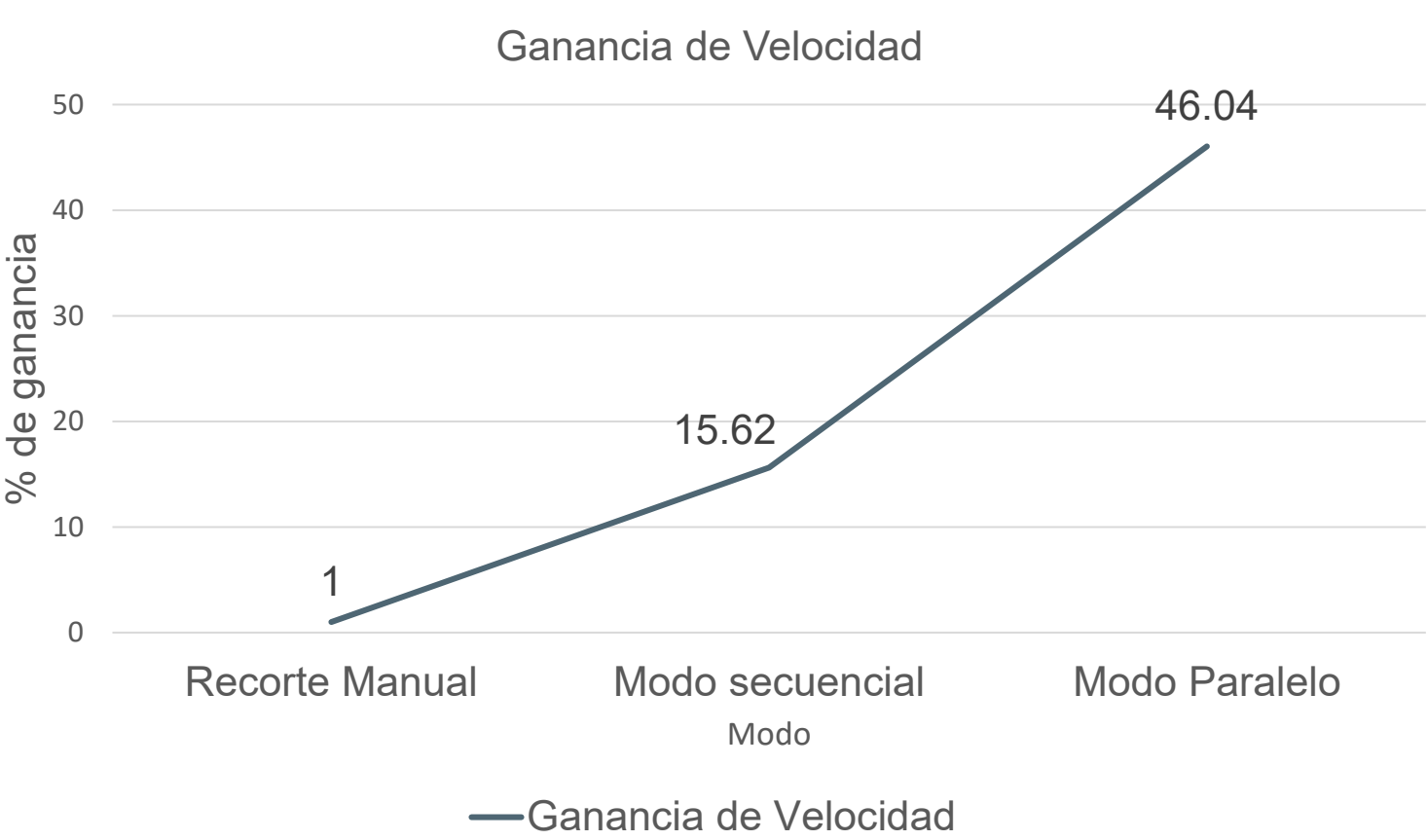
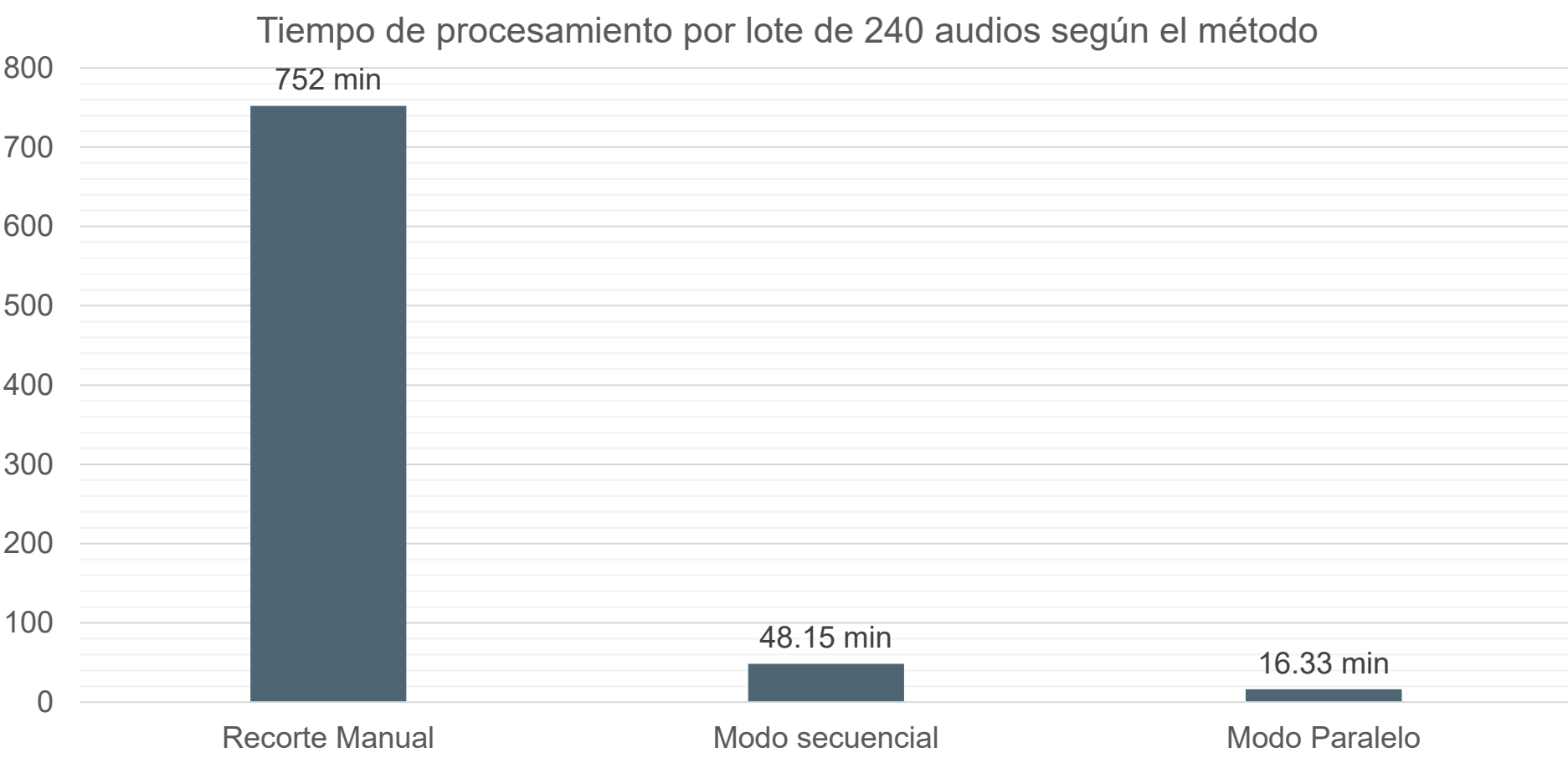


Figura 1. Esquema de preprocesamiento e inferencia en paralelo.

RESULTADOS

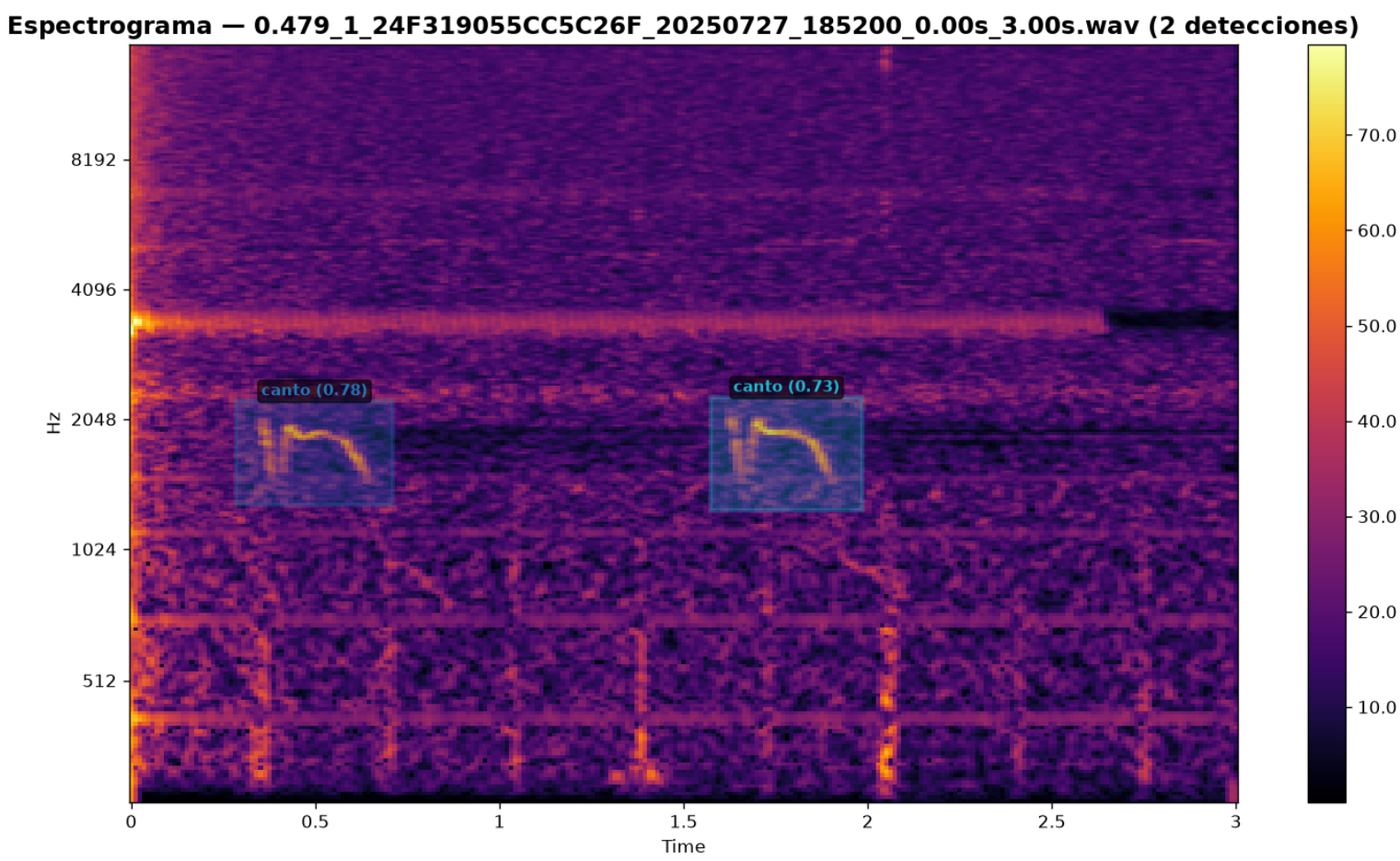
Comparativa de rendimiento en el procesamiento de audio según el método empleado



Evaluación y resultados del modelo de detección YOLO26

	Version 1	Version 2	Version 3
Datos	511	822	2060
Recall	0.643	0.641	0.746
Precision	0.619	0.772	0.826
mAP@50	0.587	0.652	0.804
mAP@50-95	0.331	0.368	0.435

Tabla 1. Evolución del desempeño del modelo YOLO según la cantidad de datos de entrenamiento.



CONCLUSIONES

- El pipeline de preprocesamiento, basado en un filtro Butterworth, PCEN y aumentación de datos con hard negatives, redujo eficazmente el ruido ambiental y mejoró la capacidad del modelo para detectar vocalizaciones en condiciones acústicas complejas.
- El modelo YOLO26 demostró un desempeño competitivo para la detección automática de vocalizaciones de aves, alcanzando una Precisión de 0.826, un Recall de 0.746 y un mAP@50 de 0.804, permitiendo localizar cantos de forma precisa en tiempo y frecuencia.
- La incorporación de procesamiento paralelo obtuvo una aceleración de hasta 46×, reduciendo aproximadamente un 97.8% el tiempo de procesamiento respecto al recorte manual, lo que hace viable el análisis masivo de grabaciones bioacústicas.