



# ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

**DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES**

**CARRERA DE INGENIERÍA EN ELECTRÓNICA Y TELECOMUNICACIONES**

**TEMA:**

**“Diseño de un clasificador para la identificación de los sonidos emitidos por aves mediante técnicas de aprendizaje no supervisado y aprendizaje profundo”**

**Elaborado por:**

Herrera Jaramillo, Estefania Alejandra

**Director del Proyecto:**

Ing. Carrera Erazo, Enrique Vinicio

Sangolquí, 2023



# AGENDA

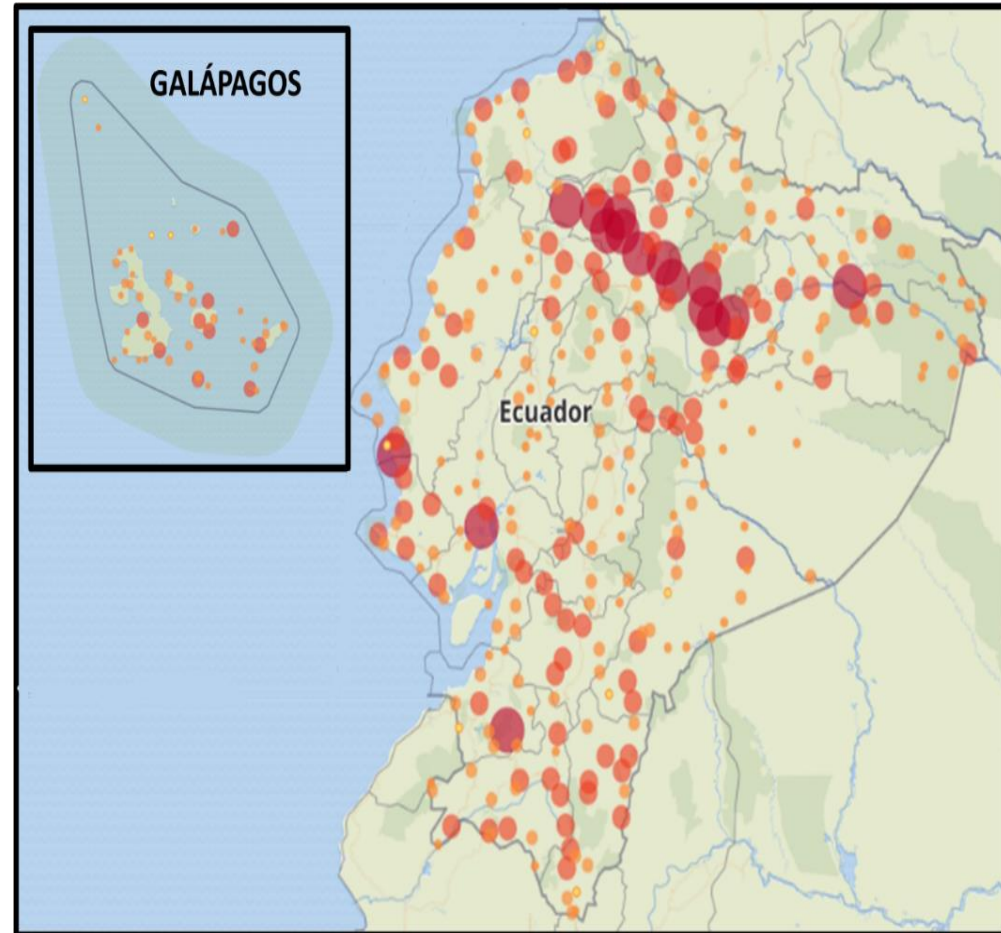
- 1 Introducción
- 2 Objetivos
- 3 Desarrollo e Implementación
- 4 Resultados
- 5 Conclusiones
- 6 Trabajos Futuros



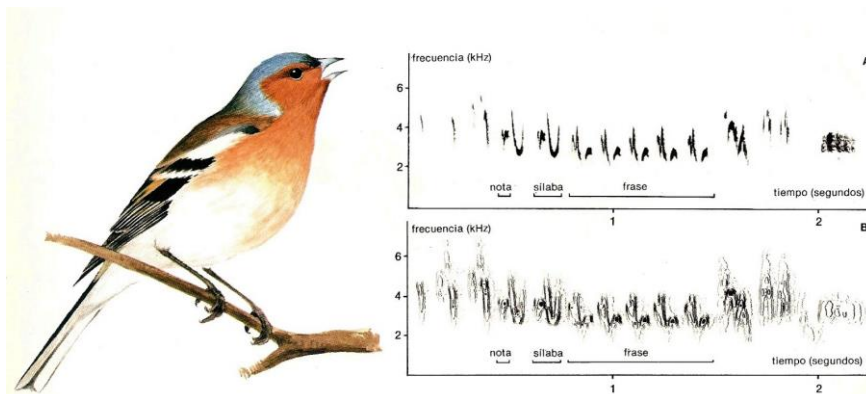
# Introducción



- Biodiversidad de aves:  
aproximadamente 1660.
- Muchas de estas especies de  
aves se encuentran en peligro  
de extinción.



- Centro de observación de aves
- Dificultades para registrar las distintas especies de Aves.



- Cánticos y llamados
- Sistemas de aprendizaje automático

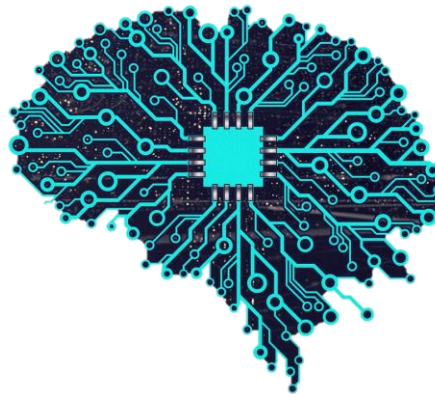


# Objetivos



## General

Diseñar un sistema que permita la clasificación de audios mediante técnicas de aprendizaje no supervisado y aprendizaje profundo para la identificación de los sonidos emitidos por las aves.



## Específicos

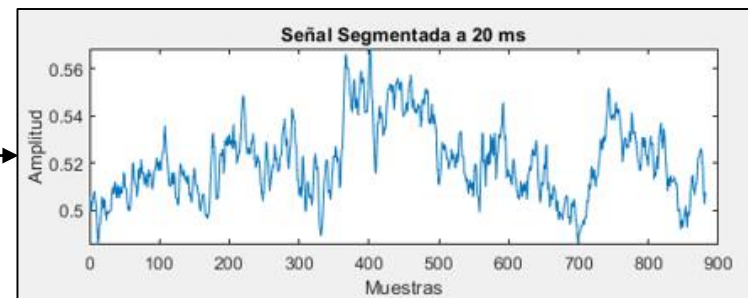
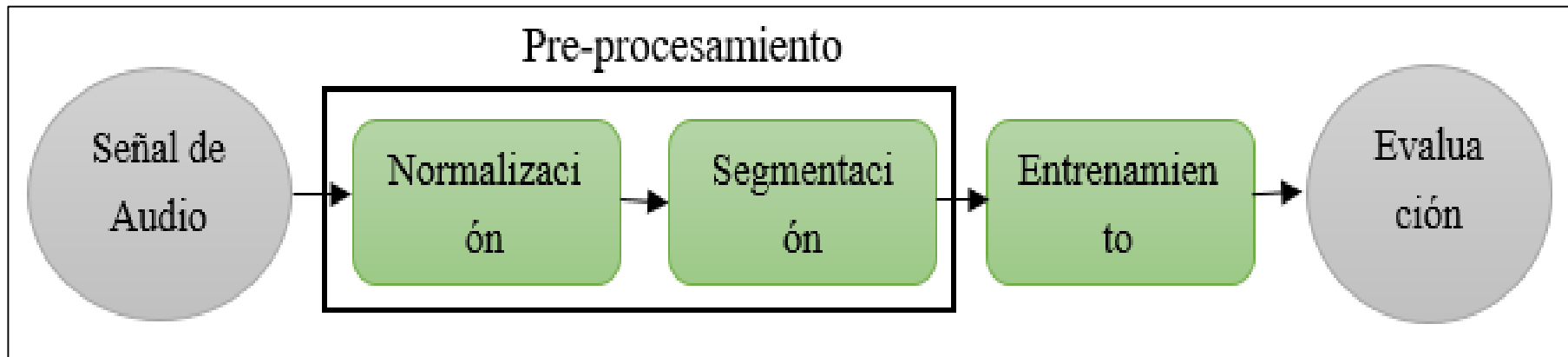
- Investigar técnicas de aprendizaje automático y aprendizaje profundo.
- Determinar una base de datos existente con sonidos de las aves.
- Implementar el modelo de autoencoder para automatizar el proceso de extracción de características y posterior clasificación.
- Comparar el modelo anterior con esquemas basados en aprendizaje profundo.
- Evaluar el sistema de reconocimiento de los sonidos.





# Desarrollo e Implementación





- Normalización a un rango entre 0 y 1 de amplitud.
- Segmentación de 20 ms, en cada segmento hay 882 muestras.
- Selección de segmentos con un valor de energía mayor.

## Entrenamiento



### Autoencoders con Softmax

- Su estructura permite replicar los datos de entrada en la salida de la red.
- El número de neuronas en la primera capa oculta es de 400 y en la segunda es de 90.
- Parámetros regularizadores
- Capa Softmax con dimensión de 200 épocas.

### Redes Neuronales LSTM

- Transformación de los datos en serie.
- Las neuronas que forman la capa oculta son reemplazadas por memorias.
- Se reduce el problema del desvanecimiento de la gradiente.

# Resultados



## Autoencoders más capa Softmax

Se realiza la evaluación del sistema bajo varios escenarios.

#Ventanas	Escenario	Exactitud (%)	Error Medio
<b>20 ms</b>	10 audios	99.2%	0.0025954
	50 audios	95.9%	0.00072511
	100 audios	84.3%	0.00026524
<b>60 ms</b>	10 audios	98.5%	0.0057439
	50 audios	95.3%	0.0015113
	100 audios	83.5%	0.0013394

Al segmentar las señales de audio con ventanas de 20 ms, la respuesta del sistema presenta un mejor resultado en la exactitud del aprendizaje de la red.



## Redes Neuronales LSTM

Se realiza la evaluación del sistema bajo varios escenarios.

Escenario	Interacciones	Tiempo	Exactitud (%)
10 audios con LPCs	400	32 sec	46.94 %
10 audios sin LPCs	400	19 min 3 sec	29.08%
60 audios y ventanas de 500 ms	5200	11 min 11 sec	74.16%
60 audios y ventanas de 900 ms	2800	5 min 49 sec	73.40%
100 audios y ventanas de 500 ms	8400	34 min 23 sec	78.75%
130 audios y ventanas de 500 ms	10400	30 min 54 sec	79.14%



## Redes Neuronales LSTM

Con las pruebas realizadas se puede evidenciar que el rendimiento aún se encuentra bajo, por lo que se realizan nuevas pruebas aumentando el orden de los coeficientes LPC, permitiendo el paso de más segmentos que puedan contener información válida para la identificación.

Escenario	Interacciones	Tiempo	Exactitud (%)
100 audios	9000	72 min 1 sec	84.49 %
130 audios	11400	100 min 21 sec	85.67%



# Conclusiones



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA



- Con la implementación del modelo propuesto de aprendizaje no supervisado con autoencoders, se logra obtener resultados aceptables. Al trabajar con señales de audio, este modelo cuenta con características que reduce la distorsión permitiendo llegar a una exactitud del 99.2%. Al utilizar un autoencoder con dos capas ocultas llevó al fortalecimiento en el entrenamiento ayudando a incrementar la exactitud, obteniendo un error medio del 0.0025954.
- Se comprendió el funcionamiento de las redes recurrentes LSTM, para una mejor predicción de la red se realiza el procesamiento previo de los datos utilizando coeficientes LPCs, con ello se obtuvo una reducción en el tiempo de procesamiento de 19 min con 3 sec a tan solo 32 sec trabajando con 10 audios, al trabajar con 130 audios y utilizando los LPCs el tiempo subió a 100 min con 21 sec, ya que se ingresó más información al proceso. La exactitud del modelo mejoró con la utilización de estos coeficientes hasta llegar a un porcentaje del 85.67%.



- Al aumentar la misma cantidad de audios a procesar en los modelos implementados, se tiene que con 100 audios ambos modelos presentan un rendimiento alrededor del 84%, las redes recurrentes LSTM llegan al nivel del rendimiento de los autoencoders. El comportamiento difiere al aumentar la cantidad de audios, por la parte de los autoencoders disminuye su rendimiento, pero las redes recurrentes LSTM incrementa el valor de la exactitud del sistema.
- Finalmente se logró implementar un clasificador para la identificación de los sonidos emitidos por las aves. Con este sistema se puede ayudar en los centros de protección de aves silvestres al poder reconocer ciertas especies que se encuentren en la zona, ya que, al ingresar una cantidad específica de audios, la probabilidad de aprendizaje será alta.



# Trabajos Futuros



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

- Puede extenderse hacia la aplicación de otras técnicas tanto de machine learning como de Aprendizaje profundo, y observar la variación en las métricas de rendimiento.
- Sería interesante aplicar este sistema para identificar el sonido de otras especies de fauna silvestre, de instrumentos musicales, de automóviles o en la identificación del idioma de las personas, que presentan audios más largos y complejos.



# Gracias



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA