

Machine Learning: Clasificación de ruido en un Audio

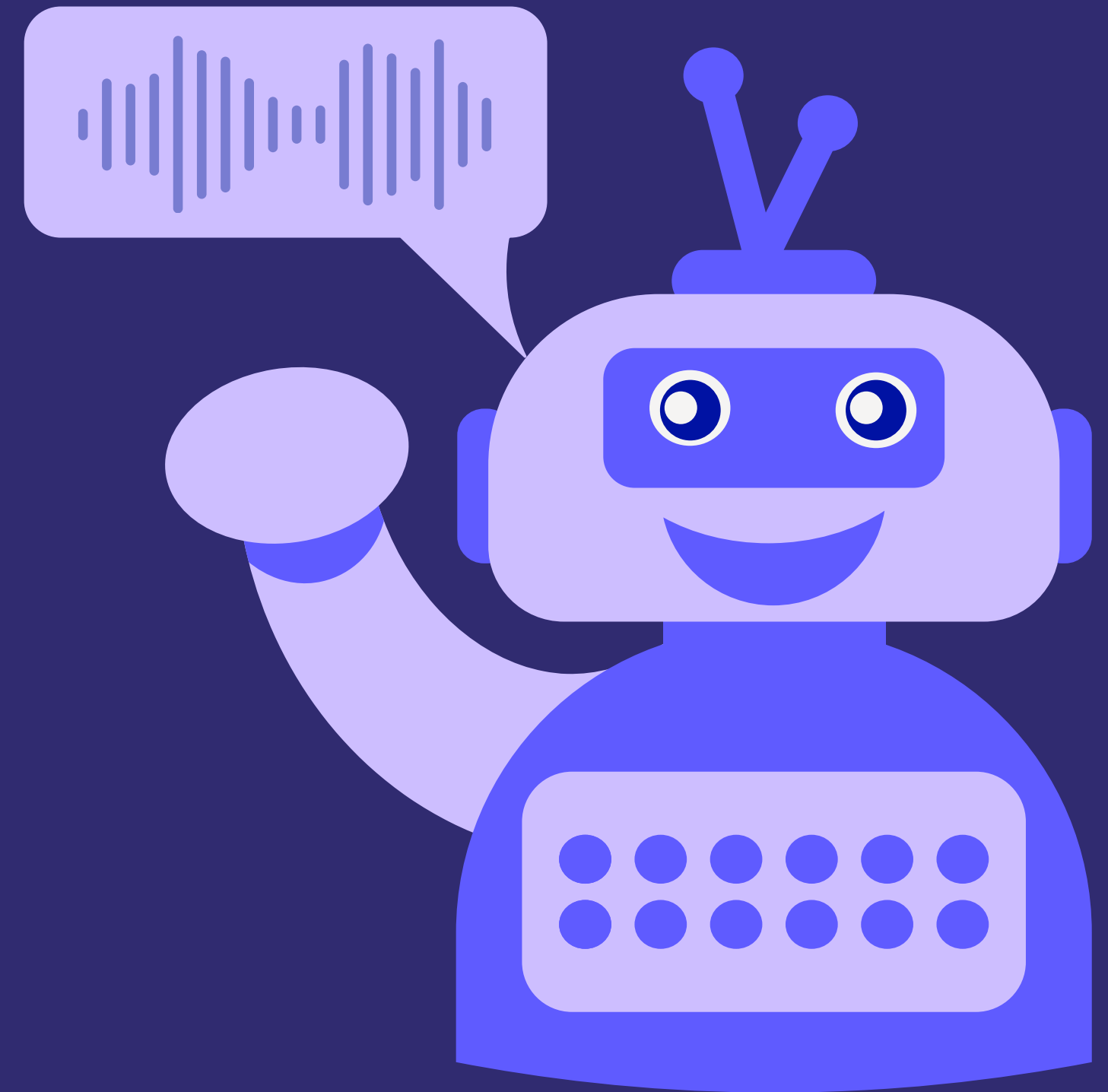
Ricardo Alexander Castillo Sandoval
Elias Obed Flores Martin
Valeria Ibarra Hernandez

Fecha 04/03/2025



Indice

Introducción.....	1
Marco teórico.....	3
Metodología.....	5
Modelo.....	7
Discusión.....	12
Conclusiones.....	13



Introducción

La calidad del audio es fundamental en aplicaciones como la industria, la investigación y la salud auditiva.

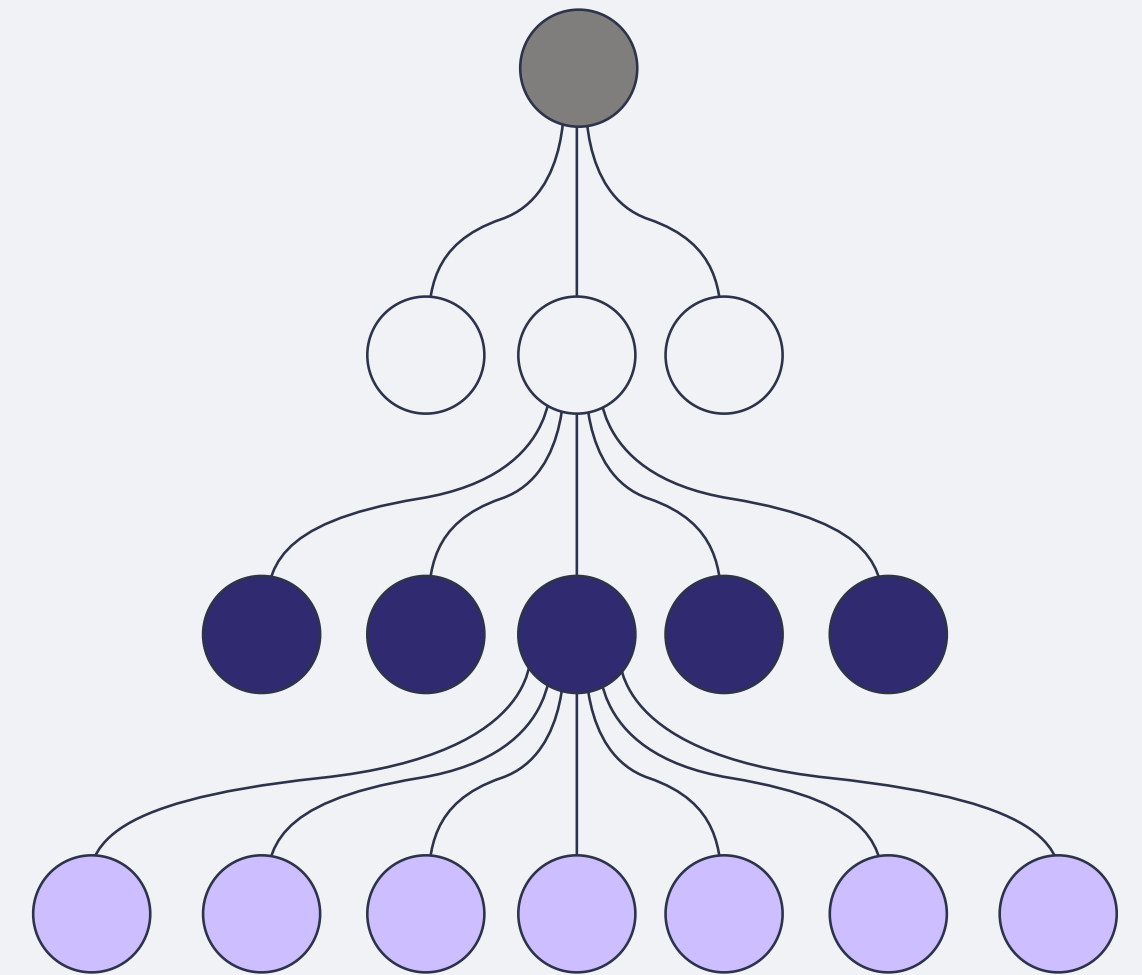
La clasificación de ruido es un paso clave en el procesamiento de audio, ya que permite mejorar su claridad y facilitar tareas que dependen de la identificación de señales ruidosas.



Por que una CNN?

Las redes neuronales convolucionales, estas han demostrado ser una alternativa superior a métodos matemáticos tradicionales.

- Eficiencia en procesamiento de audio
- Robustas ante transformaciones.
- Captura de patrones espaciales y temporales.



Revision de literatura

Modelo Aprendizaje Maquina	Exactitud%	Articulo	Dataset
KNN	55.48%	Desarrollo y análisis de clasificadores de señales de audio Fabian Aguirre Martin 2017	UrbanSound8K-5 IRMAS CMU PDA Database -16 GTZAN Music/Speech Database -17
Random forest	68.84%		
SVM	66.23%		
VGG16	74.6%	Detección de ruido de origen humano en el Parc de l'Albufera mediante técnicas de aprendizaje máquina Lanau Carrasco,Alba	Grabaciones del Parque Natural de la Albufera
AlexNet	82.8%		

Que hace diferente nuestra propuesta?

La mayoría de los proyectos analizados usan ML para:

- Identificar actividad humana
- Clasificar tipos de instrumentos

Nuestro enfoque:

- Clasificar audios según su legibilidad
- Separar audios comprensibles de los no legibles
- Distinguir ruidos ambientales

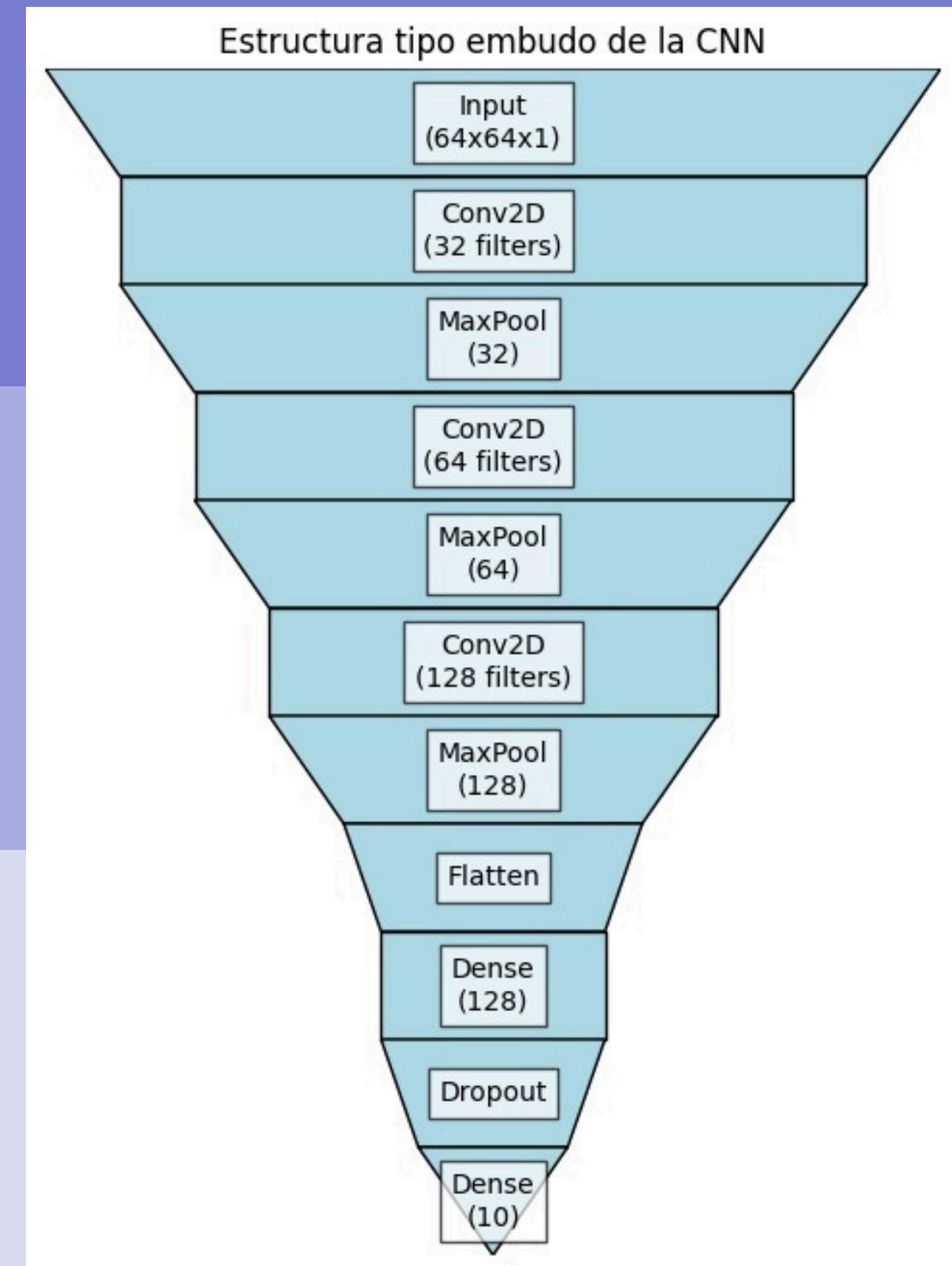


Metodologia

Arquitectura del modelo

Proceso:

1. Entrada: Imagen de 64x64 píxeles en escala de grises.
2. Capas de Convolución.
3. MaxPooling.
4. Capa de Neuronas (Fully Connected).
5. Dropout.



UrbanSound8K

Dataset de sonidos urbanos

- 8,732 clips de audio (máx. 4s) con 10 clases de sonidos urbanos.
- Formato WAV, frecuencia de muestreo 44.1 kHz.
- Grabaciones en entornos reales con ruido de fondo.

Clases de sonido:

Bocinas | Perros | Sirenas | Golpes
| Multitudes | Taladros | Metro | Aire
acondicionado | Freidoras | Motores



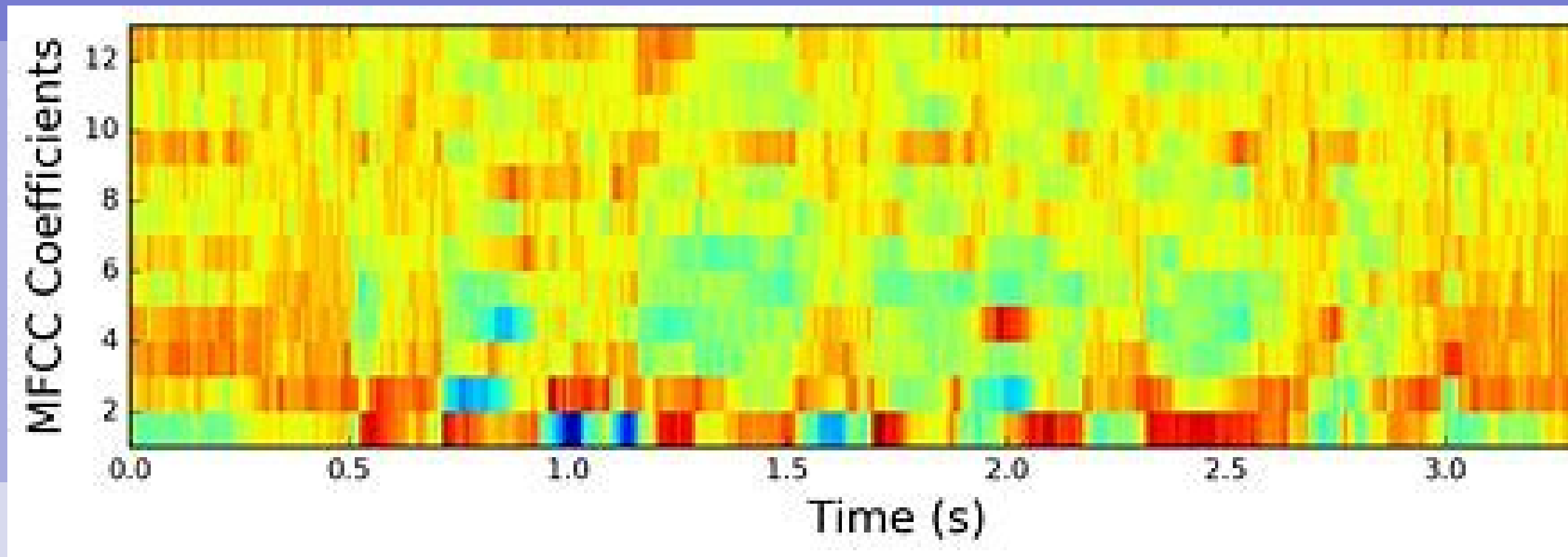
Preprocesamiento de los datos



1. Procesamiento de Datos
2. Conversión a WAV
3. Segmentación en pistas mas cortas
4. Normalización del audio
5. Extracción de características (MFCCs)
6. Normalización de características
7. Codificación de etiquetas
8. División del dataset en:
 - Entrenamiento
 - Validación
 - Prueba



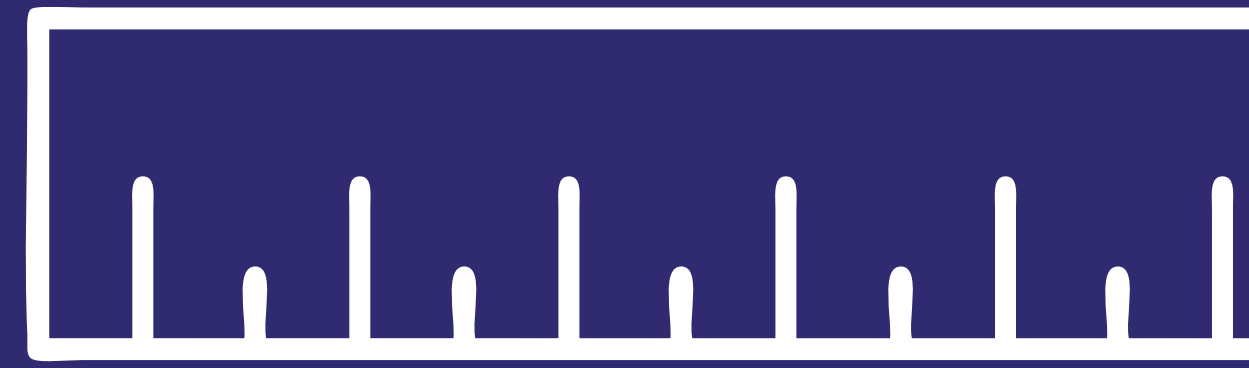
Coeficientes Cepstrales de Frecuencia Mel (MFCC)



Representación compacta de la envolvente espectral de una señal de audio, obtenida a partir de la escala Mel.



Metricas de Evaluacion



Optimizador

Optimizador ADAM

Optimizador que ajusta de manera automática las tasas de aprendizaje de los parámetros del modelo

Funcion de perdida

`categorical_crossentropy`.

Función de pérdida utilizada en redes neuronales cuando se trabaja con problemas de clasificación multiclase

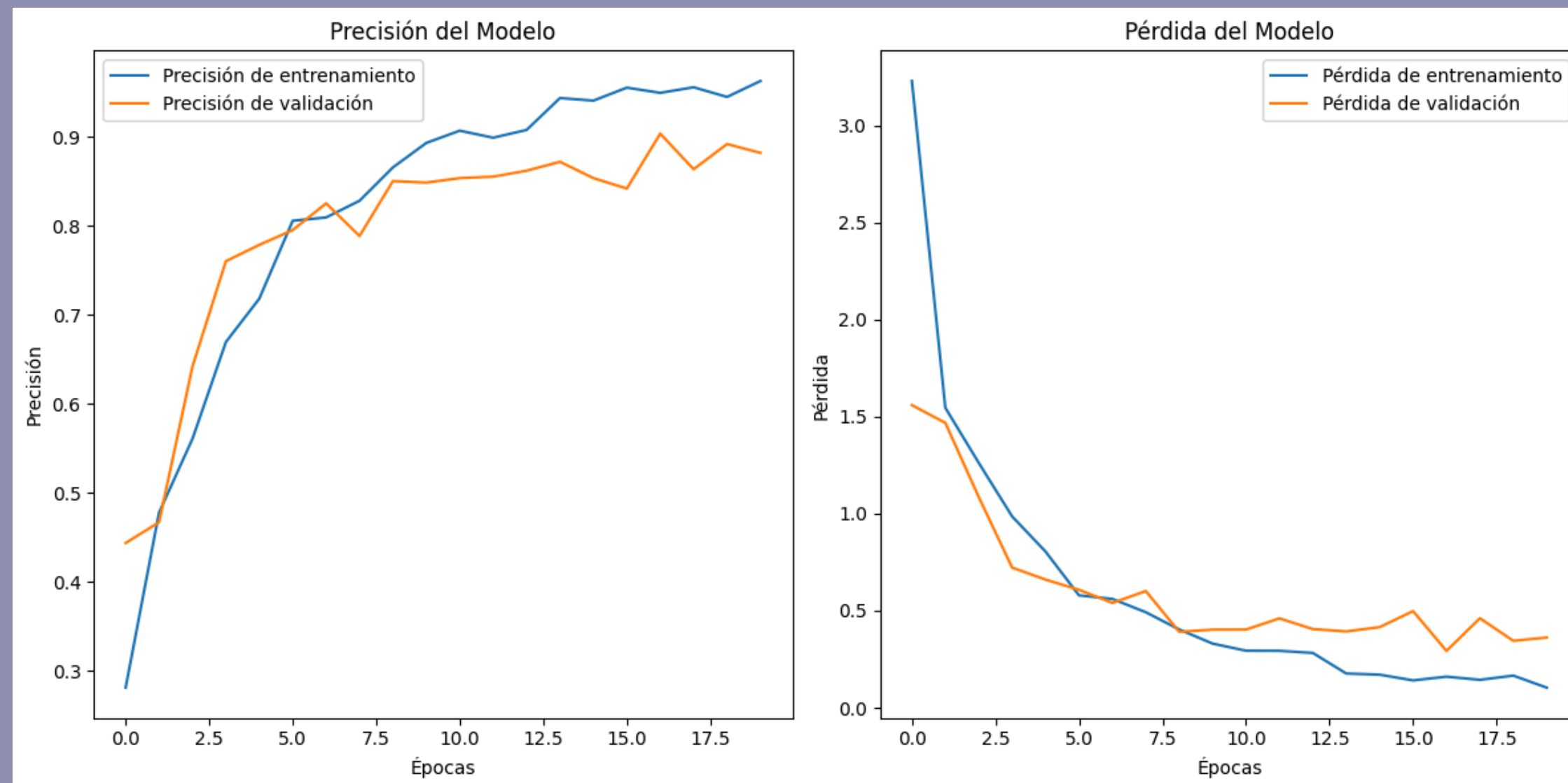
Metricas

Accuracy

accuracy es una métrica que mide qué tan bien predice tu modelo.

Evaluación del desempeño

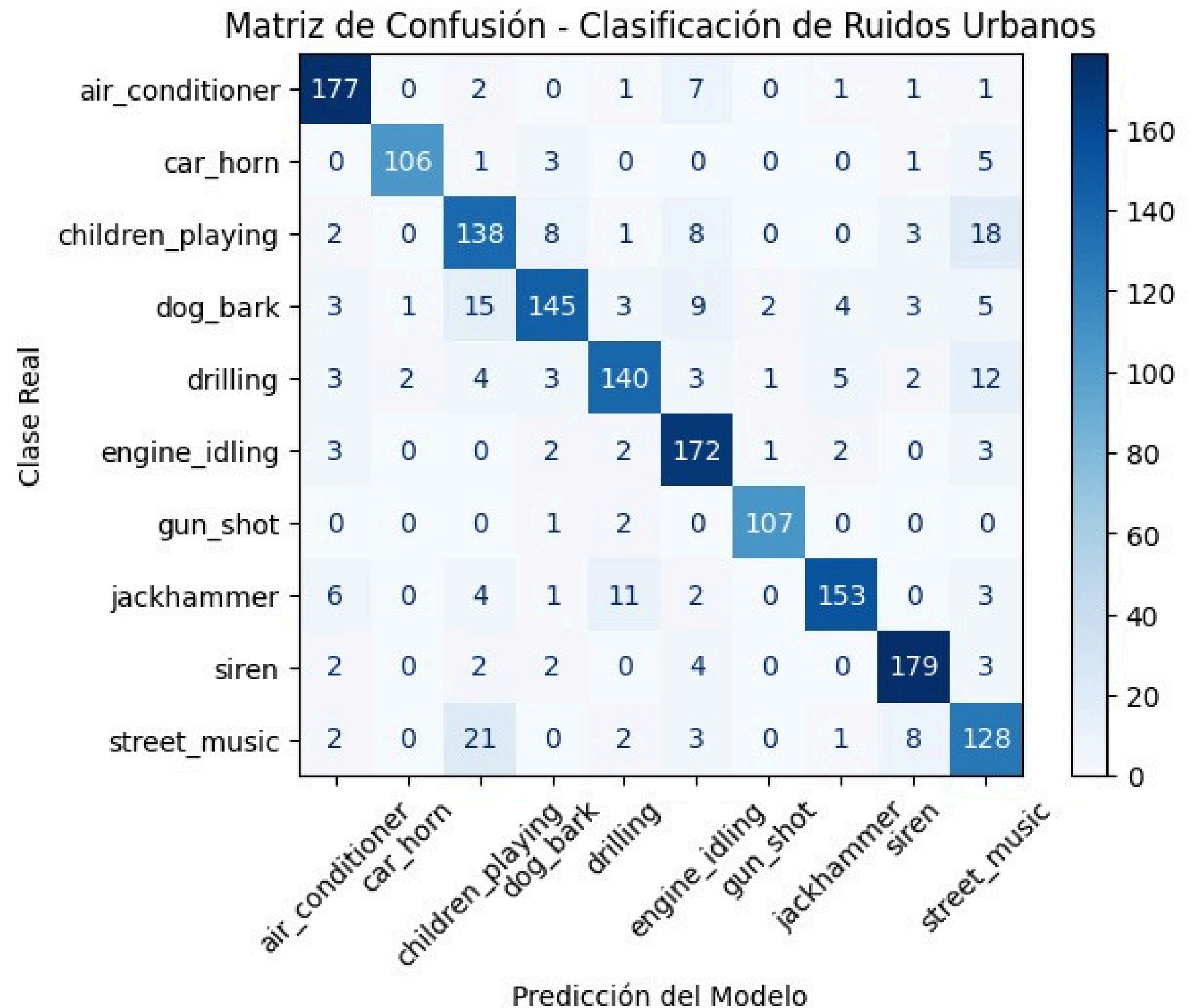
El modelo alcanza una alta precisión en los datos de entrenamiento y prueba, sin embargo, la diferencia de accuracy del entrenamiento (96.54%) y la prueba (89.81%) sugiere que puede haber cierto sobreajuste.



Matriz de confusion

Mayor complejidad con las clases :

- Children_playing
- Dog_bark
- Driling
- Street_music



• • • Discussion • • •

- El procesamiento de audio es complejo y existen múltiples formas de abordarlo.
- La extracción de características es clave.
- El rendimiento depende de:
 - Calidad del dataset
 - Herramientas de grabación
 - Tipos de ruido presentes
- Aplicaciones potenciales:
 - Mejora en videollamadas
 - Optimización de calidad de audio
 - Avances en auxiliares auditivos



Conclusion

- El procesamiento de audio por computadora tiene un gran potencial en futuros proyectos tecnológicos.
- Los resultados obtenidos del modelo fueron positivos pero con un gran margen de mejora en precisión y en funcionalidad.
-
- Sus aplicaciones incluyen el procesamiento de voz, la mejora en la nitidez de conversaciones virtuales en entornos ruidosos y optimización de auxiliares auditivos





Referencias

1. Aguirre Martín, F. (2017). Desarrollo y análisis de clasificadores de señales de audio [Tesis de grado, Universidad Politécnica de Valencia]. Repositorio UPV.
2. Lanau Carrasco, A. (2023). Detección de ruido de origen humano en el Parc de l'Albufera mediante técnicas de aprendizaje máquina [Tesis de grado, Universitat Politècnica de València]. Repositorio UPV.
3. Orozco-Reyes, L., Arévalo, M. A. A., García-Canseco, E., & Ibarra-Hernández, R. F. (2022). Clasificación de la señal de audio cardíaco mediante la transformada de Fourier de tiempo corto y aprendizaje profundo. Res. Comput. Sci., 151(7), 141-155.

