

# Clasificación automática de sonidos utilizando aprendizaje máquina

You

October 19, 2023

## Abstract

Your abstract.

## 1 RESUMEN

aprendizaje máquina para el reconocimiento de sonidos. Algunos sonidos son fácilmente distinguibles, como una risa, pero otros pueden ser muy similares entre sí, como una batidora y una motosierra. Además, la variabilidad inherente a estos audios hace que este problema sea bastante complicado de resolver mediante técnicas de procesamiento clásicas, pero supone un desafío apropiado para los altos niveles de abstracción que se pueden conseguir con las técnicas de aprendizaje máquina. El trabajo presenta dos modelos de red neuronal convolucional (CNN) para resolver un problema de clasificación de sonidos ambientales en siete categorías distintas. Los extractos de audio usados son los proporcionados por la base de datos UrbanSound8K.

## 2 OBJETIVO

El objetivo de este trabajo consiste en la realización de un sistema capaz de clasificar sonidos ambientales con una precisión aceptable. Para ello se estudio los distintos métodos de extracción de características de audio e investigación en profundidad sobre la inteligencia artificial para elegir el mejor modelo que sea capaz de resolver este problema.

## 3 Según el tipo de problema

Se puede dividir el análisis de audio en varios tipos de problemas. Se distinguen en dos grandes grupos, dependiendo si el resultado del audio a analizar va a tener información temporal o no. Entonces sería un problema de detección de audio. En el otro caso, tenemos que hacer distinción de otros dos problemas. Si la salida nos va a mostrar la clase o el tipo de sonido que se le ha introducido, estamos ante un problema de clasificación de audio. Por último, si la salida va a ser varias etiquetas del sonido, el problema será de etiquetado de audio.



Figure 1: Esquemas de los distintos sistemas de análisis de audio

## 4 DEEP LEARNING PARA ANÁLISIS DE AUDIO

El sonido es una onda de presión longitudinal formada por las compresiones y rarefacciones de las moléculas de aire, que se propaga en dirección paralela a la aplicación de energía.

### 4.1 Conversión analógico-digital:

- Filtrado paso bajo.
- Muestreo temporal.
- Cuantificación.



Figure 2: Esquema de conversión analógico-digital

## 5 Espectrograma

Para obtener una representación tiempo-frecuencia se debe hacer la DFT de la señal cada cierto tiempo. En el eje de abscisas representamos el tiempo transcurrido y en el eje de ordenadas las frecuencias.

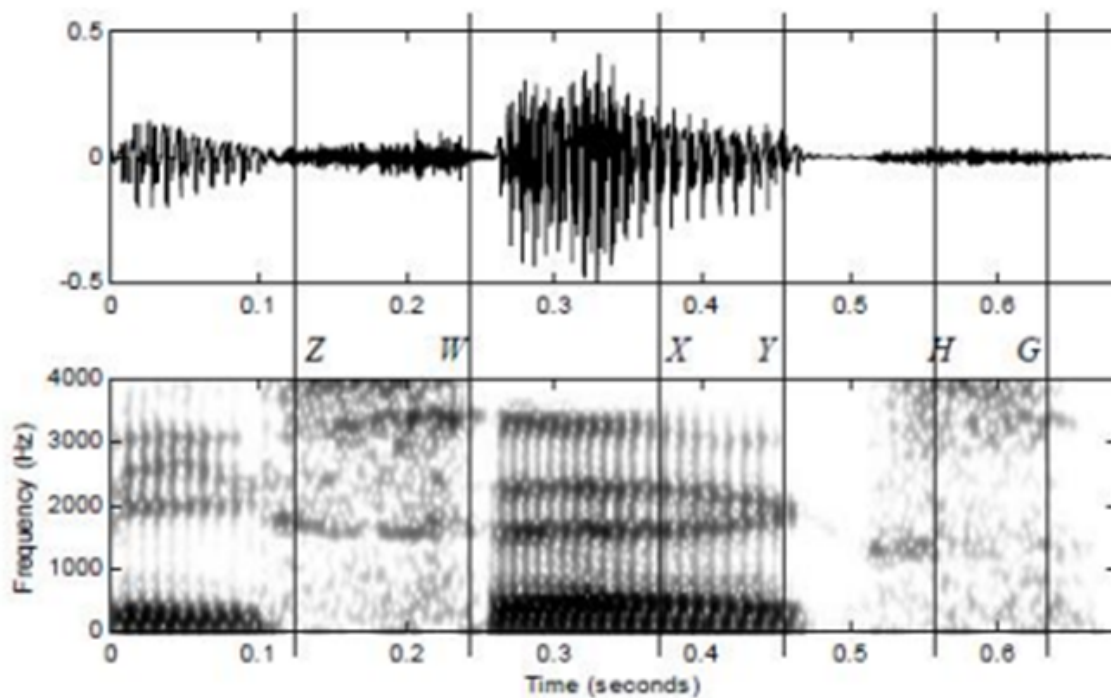


Figure 3: Esquema de conversión analógico-digital

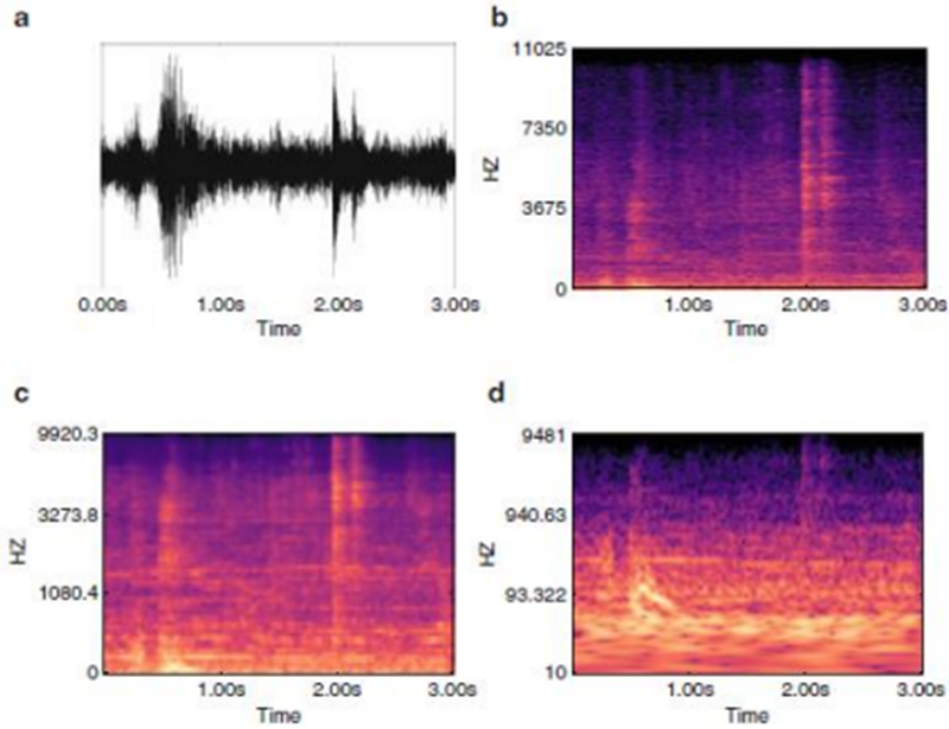


Figure 4: Distintos ejemplos de espectrogramas

## 6 metodología

- Se uso una base de datos proporcionada por Salamon et al. llamada UrbanSound8K.
- Esta dataset está compuesto por 8732 extractos de audio en formato WAV, con una duración no superior a 4 segundos. Estos extractos de audio están sacados de la página web de FreeSound y clasificados en 10 categorías: aire acondicionado, claxon de coche, niños jugando, ladrido de perro, máquina taladradora, motor al ralentí, disparo de arma, martillo neumático, sirena y música sonando en la calle.

Número identificador	Clase	Número identificador	Clase
0	<i>air_conditioner</i>	5	<i>engine_idling</i>
1	<i>car_horn</i>	6	<i>gun_shot</i>
2	<i>children_playing</i>	7	<i>jackhammer</i>
3	<i>dog_bark</i>	8	<i>siren</i>
4	<i>drilling</i>	9	<i>street_music</i>

Figure 5: Identificadores de clases de la base de datos

## 7 RED NEURONAL

Para conformar la red necesitamos una capa de entrada, que recibirá los datos de entrada del modelo, una serie de capas intermedias que ajustarán sus pesos para adaptarse a dichos datos, y una capa de salida, que proveerá el resultado de la red

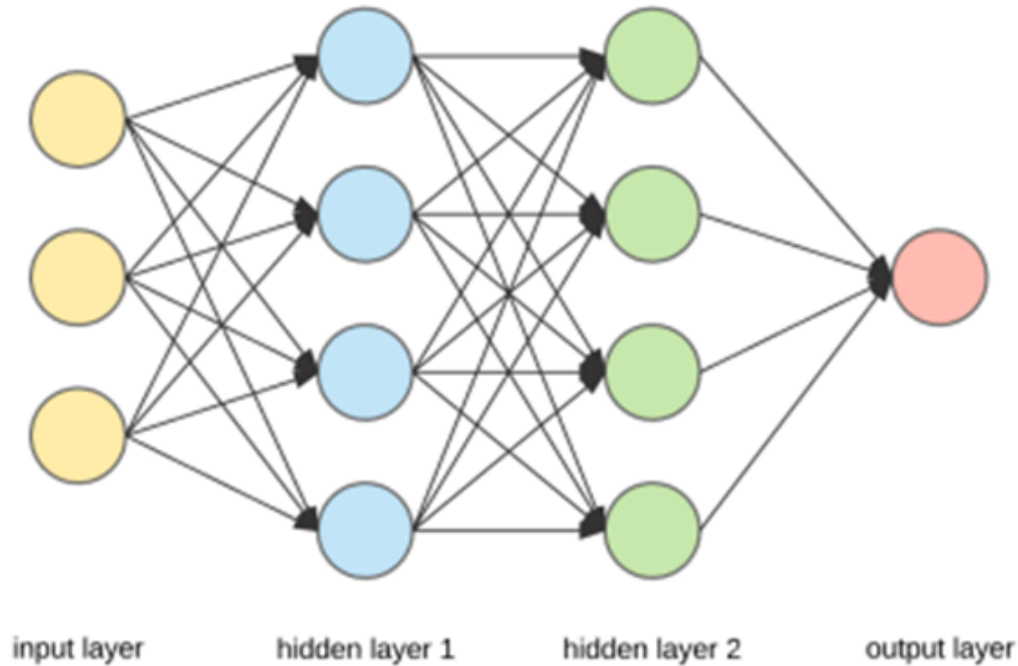


Figure 6: Esquema de una red neuronal artificial

## 8 conclusión

La clasificación automática de sonidos mediante el aprendizaje máquina representa una herramienta eficiente con diversas aplicaciones, desde la seguridad y la salud hasta la industria y el entretenimiento, permitiendo el procesamiento ágil de grandes conjuntos de datos de audio, aunque presenta desafíos técnicos, como la variabilidad de los datos y la selección de algoritmos adecuados, pero con los avances en aprendizaje profundo, como las redes neuronales convolucionales y recurrentes, esta tecnología promete mejorar la interacción entre humanos y máquinas, permitiendo una comprensión más sofisticada y respuestas más efectivas a eventos auditivos.