

## Resumen

En el marco del estudio de la teoría musical, el concepto de género ha sido definido desde distintas perspectivas.

Este trabajo analiza diversos métodos dentro del campo del aprendizaje automatizado buscando en el proceso descubrir cual es aquel modelo que daba un mejor resultado a la hora de distinguir entre las diferentes piezas de audio.

También se analizó cuáles eran las formas más óptimas de representación a la hora de realizar una clasificación por género.

## Introducción

El siguiente proyecto busca explorar por medio de diferentes algoritmos cual es el aquel más adecuado para dicha tarea, esto con el fin de aumentar la eficiencia en la clasificación de bases de datos musicales. Pudiendo ayudar a los sistemas de recomendación de música y a los generadores de listas de reproducción.

A lo largo del desarrollo de este proyecto se han encontrado 2 desafíos principales:

- Los géneros musicales están vagamente definidos, hasta el punto en que son comunes las discusiones sobre el género de una canción.
- No es una tarea sencilla la extracción de características al no estar claramente definidas en los datos del audio a la hora de entrar en el modelo

Con el fin de superar estos desafíos se ha optado por el Deep Learning como una de las opciones más adecuadas por su potencial para encontrar patrones implícitos por medio del entrenamiento, predicción y calificación de aquellos datos que el sistema reconozca como claves a la hora de cumplir con la tarea.

## Proceso y método

A continuación se hablará de los datos, técnicas y procedimientos realizados para lograr este esquema planteado:

- Se utilizan tanto audios como imágenes espectrales de estos para lograr una clasificación de los géneros musicales en cuestión.
- Los géneros fueron: Jazz, clásica, hip hop, reggae, blues, disco, country, metal, pop y rock.
- Se trabajó con diferentes métodos de deep learning como: DNN, CNN, RNN y una muestra de Autoencoders.
- Entorno : Se utilizó google colab para el desarrollo computacional.

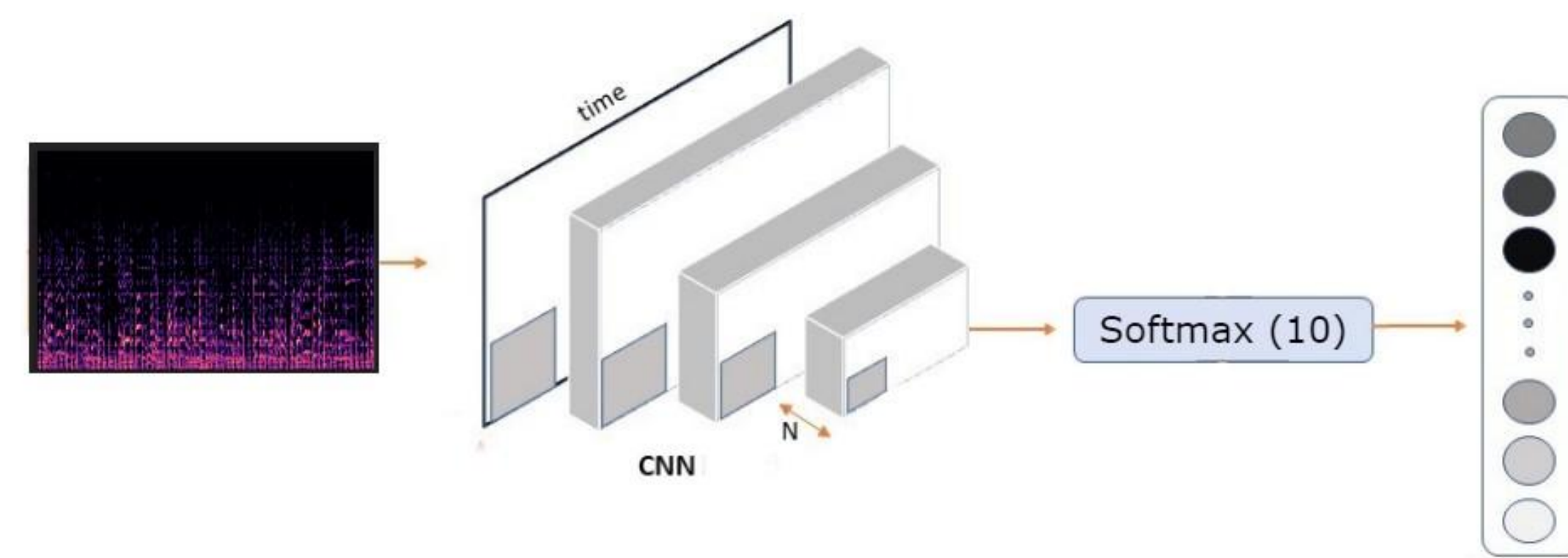


Figura 1. Modelo convolucional

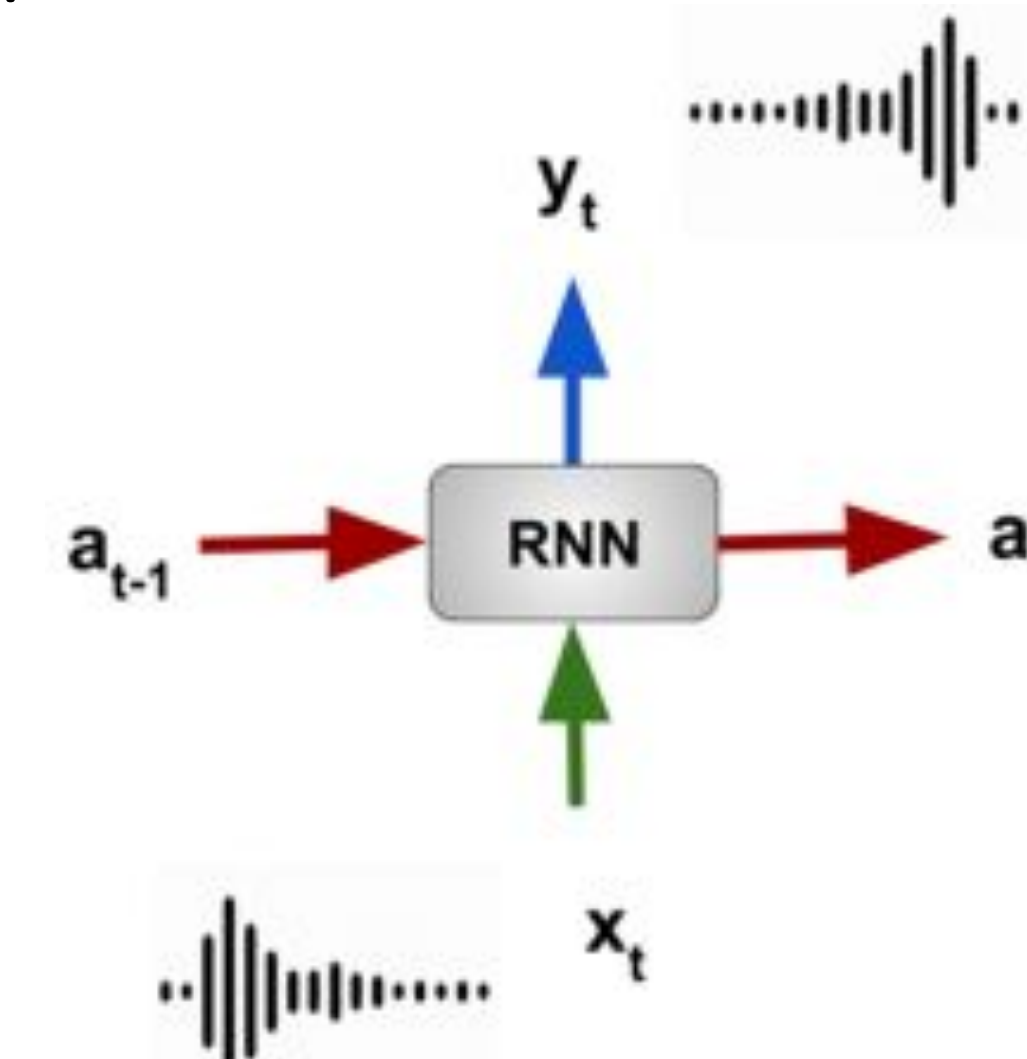


Figura 2. Red neuronal recurrente

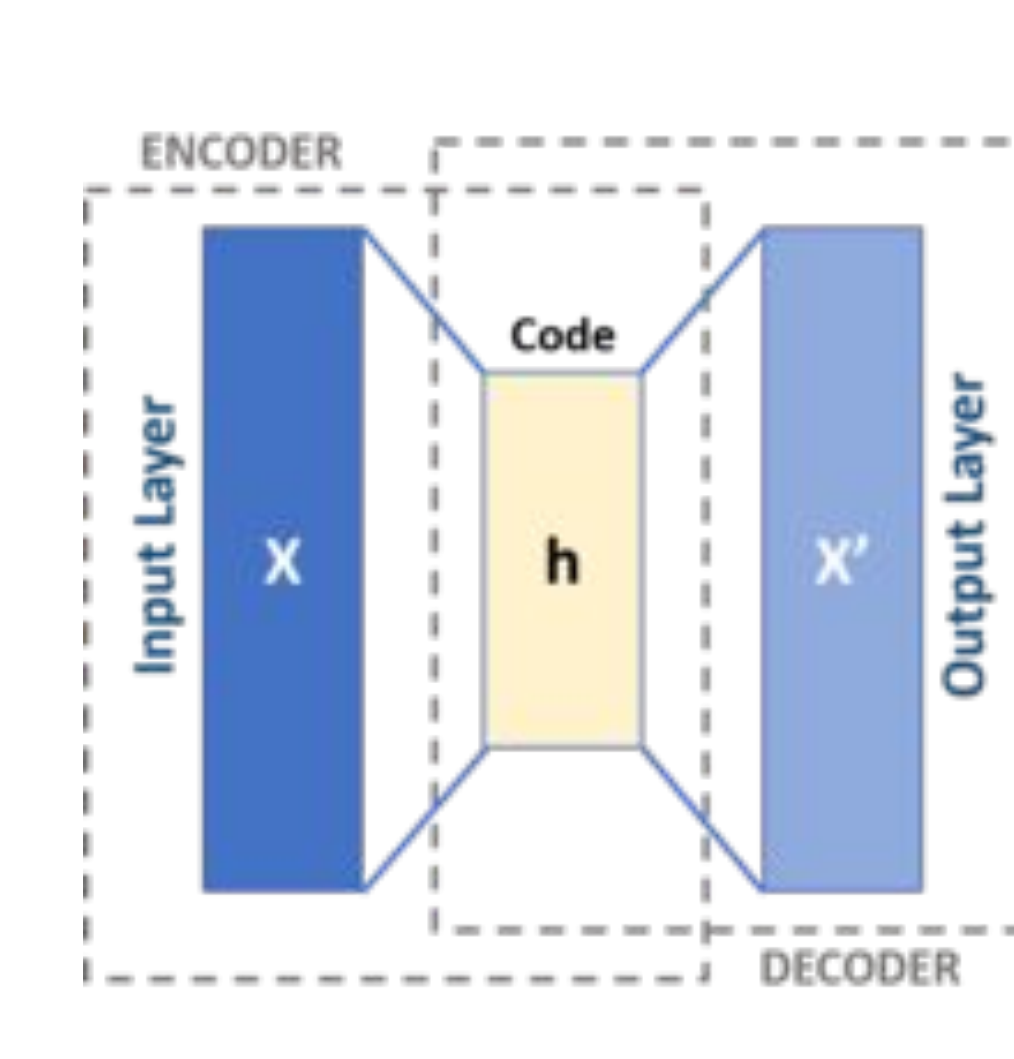


Figura 3. Autoencoder

## Conclusiones

Para este proyecto funcionó mucho mejor la implementación de una RNN para la clasificación de géneros aunque no se haya obtenido la precisión esperada y esto puede deberse a que para este tipo de problemas la “memoria” es bastante relevante.

La música tiene un ritmo y una secuencia que tiene que estar muy bien concatenada para que sea amigable al oído y cada género tiene su “ítem” que lo hace especial y diferente.

## Trabajo Futuro

Este es un proyecto en marcha y se deja como consejo que el siguiente paso a seguir, sea la implementación de una red convolucional (CNN) embebida a una RNN, existen antecedentes de esto y quizá mejore la precisión con la cual se clasifica.

## Resultados

Se evidenció lo desafiante que es trabajar con datos de secuencia (música) y lo importante que es aprender ciertos patrones específicos de cada género musical, y más aún, que una red sea capaz de diferenciarlos y poder clasificarlos de acuerdo a eso.

Iniciando con el desarrollo del proyecto se implementó una red neuronal densa como un plan piloto para conocer cómo tratar los datos obtenidos y como proceder a utilizarlos de la mejor manera. Finalizando este proceso, se procedió a aprovechar del marco teórico existente y se utilizó una red neuronal convolucional (CNN) pre entrenada a la cual se le hicieron ciertos ajustes de acuerdo a los datos existentes. Como se puede evidenciar en el *gráfico 1*, arrojó buenos resultados, aunque no el mejor.

Terminado con este método, se consiguió hacer uso de una red neuronal recurrente (RNN) la cual guarda los datos que pasan por ella y se simula una “memoria” que nos fue de mucha ayuda para el objetivo de este proyecto. Como se observa en el *gráfico 2*, este fue el mejor método implementado, superando por muy poco a lo que se denomina como Transfer Learning, que sería el método anteriormente mencionado. Por último, analizando el *gráfico 3* nos percatamos de una comparación de las ondas de sonido entre el audio del dataset y el audio generado por el autoencoder propuesto, vemos una diferencia, tal vez un poco leve, pero tiene sus puntos de comparación.

Gráfico 1. Precisión de clasificación (CNN).

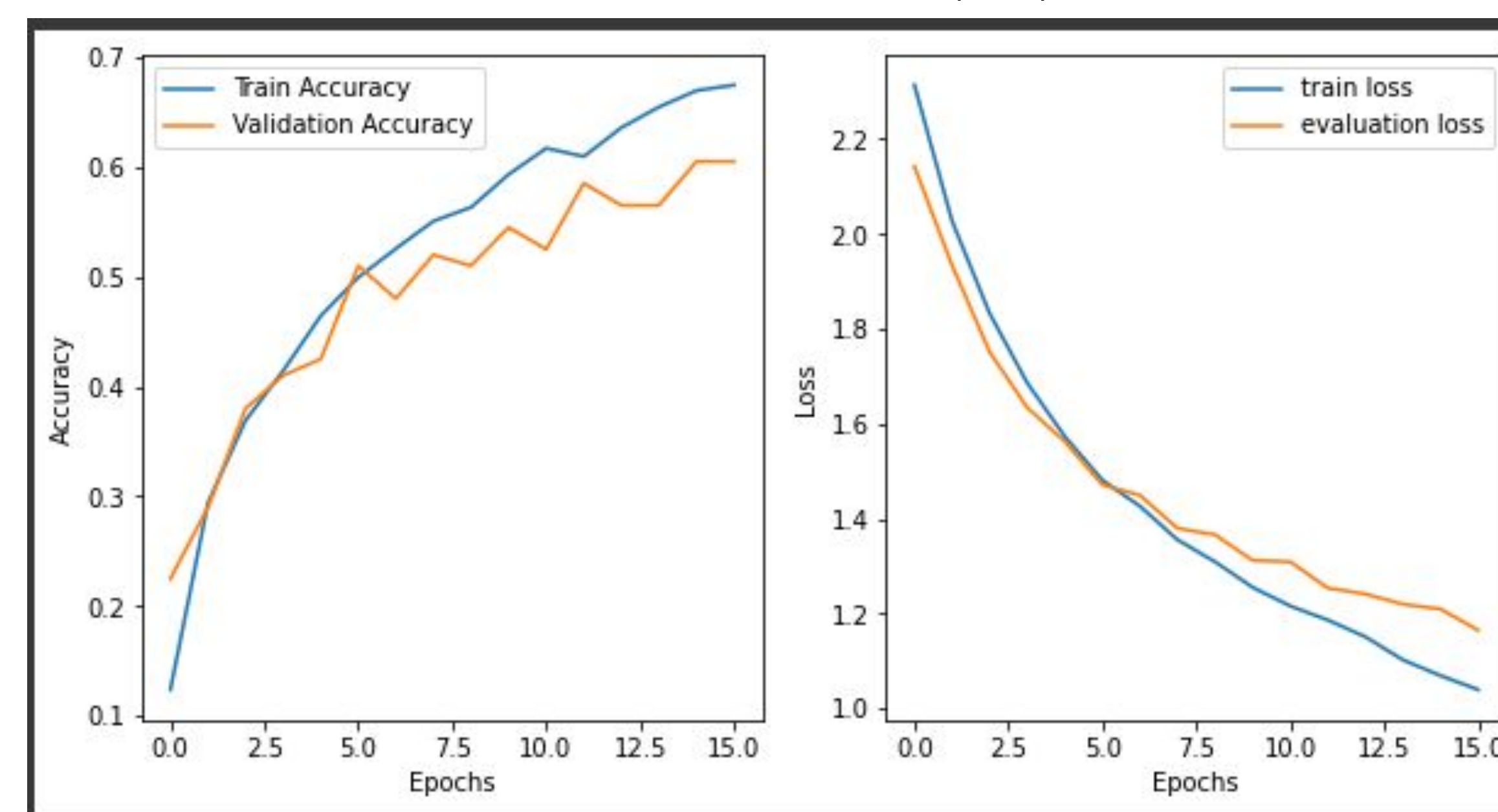


Gráfico 2. Precisión de clasificación (RNN)

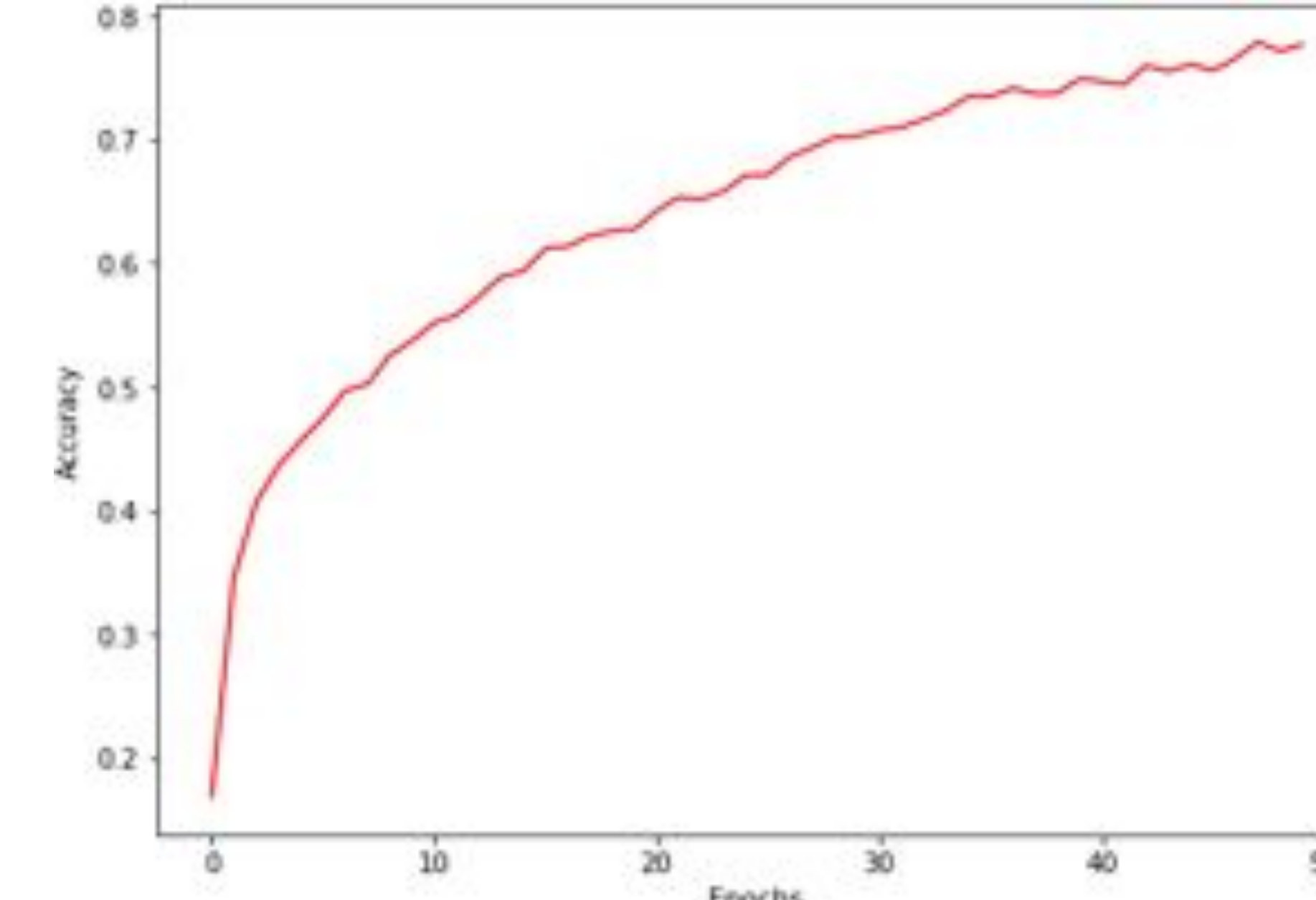
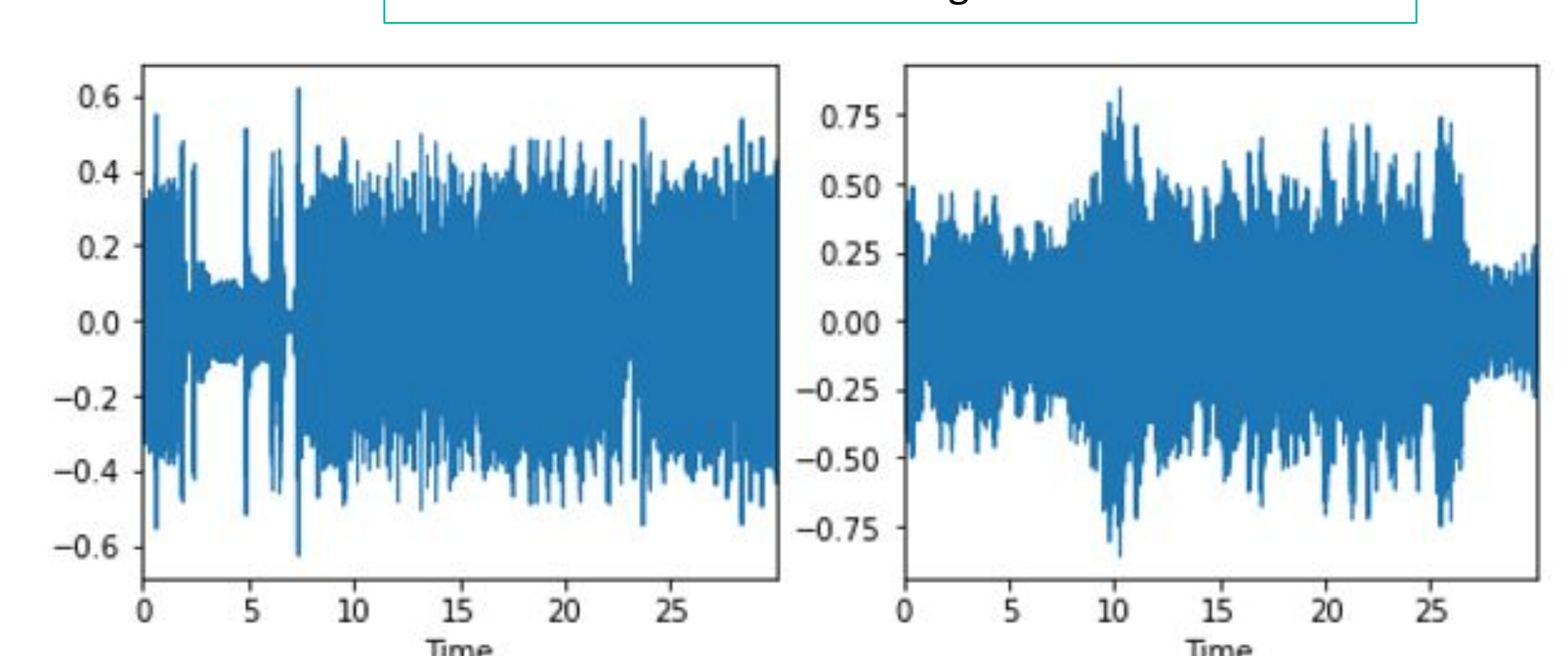


Gráfico 3. Audio real vs audio generado



## Información de contacto

Andrés Felipe Uribe García, Email: andresfelipeuribe11@gmail.com  
Juan Felipe Ortiz Trillos, Email: feltrio8@gmail.com  
Orlando Alberto Moncada Rodríguez, Email: orlandomocada610@gmail.com

## Bibliografía

1. Oramas, S. & Barbieri, F. & Nieto, O. & Serra, X. (2018). Multimodal Deep Learning for music genre classification. *J*(1), pp. 4-21.
2. (2021). Esquema de un autoencoder básico (Imagen). <https://hmlong.es/wiki/Autoencoder>